MORAL MOTIVATION:

A SURVEY OF ATTEMPTS TO UNDERSTAND THE MOTIVATIONAL QUALITIES OF MORAL JUDGMENTS

by

David Hill

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

Master of Arts in Interdisciplinary Humanities, Philosophy Stream

We accept this thesis as conforming to the required standard

.....

Dr. Myron Penner, Ph.D.; Thesis Supervisor

.....

Dr. Phillip Wiebe, Ph.D.; Second Reader

.....

Dr. Bob Doede, Ph.D.; Third Reader

CONTENTS

A. Sufficiency

B. Necessity

C. Psychopathy Unpacked

IV. Conclusion

Chapter 2: ALLAN GIBBARD'S NORM-EXPRESSIVISM

I. Norm Expressivism

A. Normative Judgment as Rationality

i. A Naturalistic Framework for Norms

ii. Psychology of Rationality

B. Emotions and Rationality

i. Voluntary and Involuntary Emotions

C. Summary of the Theory

i. Relation to Ethics

D. Review

II. Evaluation of Norm-Expressivism

A. R.M. Hare and the Logic of Normative Sentences

i. Relation to Norm-Expressivism

B. The Frege-Geach Problem

i. Gibbard's Response to the Frege-Geach Problem

C. Problems for Norm-Expressivism

i. Is Norm-Expressivism an Expressivist Theory?

ii. Normative Governance

III. Conclusion

Chapter 3: NEO-EXPRESSIVISM AND COGNITIVE THEORIES

I. Neo-Expressivism

A. Critique of Expressivism

i. Ethical Sentences Embed Motivational States

ii. A-Expression and S-Expression

B. Neo-Expressivism

i. Adapted to Motivational Internalism

ii. Propriety Conditions

iii. Summary

II. Cognitivism

A. *de jure* Sameness of Thought

B. Normative Beliefs

C. Concept Identity and Reference

D. A Cognitive Account of Motivation

III. Conclusion

Introduction

I. THE PHENOMENON OF MORAL MOTIVATION

Exposure to the moral life is common within human experience. Human beings make judgments about right and wrong actions, usually with the aim of pursuing the former and avoiding the latter. Moral judgments motivate us to act in accordance to the *moral right* and avoiding acting on *moral wrongs*. Smith makes a moral judgment that telling the truth is the right thing to do in most (if not all) situations. Smith's judgment provides him with motivation to tell the truth, even when doing so puts him at fault. This example attempts to capture the common-sense notion that we expect people to have a desire to act in accordance with what they consider to be morally right. Naturally, we are morally imperfect. People sometimes fail to do what is right because of conflicting emotions, competing desires, or apathy. Although there are times when people's moral motivations fail to compel action, it would be a mistake to say that motivation to act on their moral judgments does not exist in these cases.

Consider, for example the moral action of charitable work. If one judges this action to be right, then other motivations follow: for example, setting aside a portion of one's income to give to charitable organizations, researching different charities to see which ones provide the greatest impact, volunteering one's own time, and so on. Sometimes circumstances interfere with moral motivations. Many factors might bring about a situation where one fails to act on one's moral motivations, yet this failure does not preclude motivation. Using the example of charitable giving, I might see a World Vision commercial on TV, judge that it is right to help those in need, and yet do nothing. Despite this, my lack of action does not preclude the motivation to act on my moral judgment; in this case other factors prove stronger than my moral motivation; laziness, other financial commitments, etc. When one judges actions to be *wrong* a different kind of

motivation occurs. Consider torture, an action commonly judged to be morally wrong. People expect those who judge torture to be wrong to avoid committing torture, and also to experience rational or emotional aversion when coming into epistemic or actual contact with it.

Moral motivation is a specific example of a broader phenomenon – *normative motivation* – as other normative judgments we make also have motivating power. When acting on judgments based on *reasons* or due to the fact that we believe a course of action is *good for us*, we are motivated normatively. Normativity has to do with the "oughtness" or "shouldness" of actions. Many philosophers hold that the phenomenon of motivation is the feature which distinguishes normative judgments from non-normative judgments.[1] In contrast, judgments based on mathematics, for example that 2+2=4, or empirical data, (that a water molecule is composed of two hydrogen and one oxygen atom) do not seem to have any motivating aspects to them.

Outside of ethics, normative judgments about one's favourite foods, or what style of clothing one prefers to wear produce motivations to eat one's favourite foods, and to dress in an appealing style. Although motivation arises from many similar judgments, moral motivation functions uniquely in two ways. Firstly, one does not generally expect others to adopt motivations arising from one's own personal non-moral normative judgments. That is to say, if Phil's favourite colour is green, Phil does not expect everyone else's favourite colour to also be green. Secondly, non-moral normative judgments usually never fail to obtain unless one's environment interferes. If Dave dislikes mushrooms, that is, if Dave has made a normative

[1] "Moral Motivation," Stanford Encyclopedia of Philosophy, last modified October 19, 2006, accessed July 20, 2015, http://plato.stanford.edu/entires/moral-motivation/.

judgment that mushrooms taste bad, then Dave won't eat mushrooms.[2] In contrast, moral normative judgments often conflict with our base desires. Andrew might get a big job promotion and pay raise if he removes another candidate from consideration by slandering her, but in this case, Andrew's moral motivations are strong enough to keep him from doing so. Of course, it is not hard to imagine situations when someone *would* slander someone in order to get a job promotion, and despite the motivation to do what is right, we often ignore it. But why? This will require closer examination.[3]

It should be clearer by this point how the term 'moral motivation' is used in this paper. I use this term to refer to a person's experience of feeling *moved* to act due to their judgment that something is right. What this paper will explore are the theories philosophers provide to explain where this motivational drive originates from. To do so, moral philosophers explicate the semantics and foundations of moral judgments and motivation by dividing into three streams of thought: realist, anti-realist, and more recently, hybrid theories.

II. STATEMENT OF THE ISSUE

Moral realism regards moral facts as representing actual states of the world. As such, when people talk about generosity being "good", or pettiness being "bad", they are describing actual features of generosity and pettiness in the same way people talk about actual features of snow being, for example, white and cold. As a result, one is able to be *correct* or *incorrect* when

[2] If at a dinner where Dave is offered mushrooms and it would be rude to refuse them, Dave might then choose to eat the mushrooms. However, I contend that this sort of situation would be a moral judgment to choose not to give offense, rather than a non-moral normative judgment which failed to obtain.

[3] One well known example of this would be the Milgram experiment, where Yale University students were told by an authority figure to administer increasing electric shocks to a victim. Many participants ended up delivering electric shocks high enough to cause fatality despite the fact that doing so conflicted with their standard ethical views.

ascribing moral values to events, and, by extension, can be *correct* or *incorrect* when making moral judgments. On realism, moral motivation can be understood by appealing to the nature of moral properties themselves. This account of motivation explains that concepts like *rightness* and *goodness* intrinsically motivate without additional requirements. When one makes a moral judgment of the *good*, the judgment itself is sufficient to motivate.[4] [5] Realism is an attractive theory because of its common-sense approach to moral semantics and experience. When I say "Stealing is wrong", I use the same semantic form as when I say "The sky is blue". People are used to sentences of this form representing the world in an objective, mind-independent way which moral realism allows for.

Generally speaking, moral anti-realism or non-realism holds that moral properties and values do not exist mind-independently. On anti-realism, either moral properties and values do not exist at all, or they exist only in a mind-dependent sense.[6] Anti-realism attempts to sidestep some of the metaphysical problems faced by realists. If stealing is wrong, then what kind of a property is wrongness? Does wrongness somehow supervene on the physical properties involved in stealing and if so, how? Anti-realists say that these difficulties are unable to be answered satisfactorily by realists and so they generally characterize moral semantics as being mind-dependent expressions of approval or disapproval towards an action. Because there are no moral properties (in the realist sense) to motivate right action, anti-realists consider factors such

[4] J.D. Mackie, *Ethics: Inventing Right and Wrong* (London: Pelican Books, 1977), 23-24.

[5] When Mackie critiques Plato's view of moral motivation, he describes it has having "overwhelming" motivating force. For this paper, it is sufficient to say that moral judgments elicit a motivating force, but this force need not be overwhelming.

[6] "Moral Anti-Realism," Stanford Encyclopedia of Philosophy, last modified February 11, 2015, accessed July 20, 2015, http://plato.stanford.edu/entries/moral-anti-realism/.

as emotions, intuition, and social groups (to name a few) as plausible candidates for eliciting a motivational response to moral judgments.

Hybrid views, as the name suggests, combine relevant aspects of both realism and anti-realism. In short, hybrid theorists believe both that realist theories have difficulty satisfying empirical queries about the nature of moral properties, and that anti-realist theories fail to address the common-sense semantic notions of universality and truth moral properties seem to possess. Thus, hybrid theories seek to combine the relative strengths of both realism and anti-realism while seeking to avoid their perceived weaknesses. Generally speaking, hybrid theories focus on moral semantics with respect to realism in order to answer the question "What does it mean to say 'Lying is wrong'?" With respect to anti-realist views, there are a variety of possible arguments. As a result, hybrid moral theories that view emotion as an integral part of moral motivation look quite different from theories that focus on reasoned consideration.

This thesis will investigate the claim that anti-realist and hybrid theories provide a more robust and comprehensive description of moral motivation than realism does. The strength of moral realism is that it resonates strongly both with people's common-sense understanding of semantics and experience of perceived moral facts. By contrast, anti-realist theories claim that realists are unable to provide a satisfactory description of the precise nature of moral properties. If realists cannot provide such an answer, then it could be that moral properties are mind-dependent phenomena, based in cognitive structures which can be empirically investigated. Hybrid theorists, while borrowing heavily from anti-realist theories, charge anti-realists with not having a satisfactory semantic model of moral values. Hybrid theories are based on combinations of realist and anti-realist ideas, hoping to avoid the specific weaknesses of each.

The general question for this thesis will be, "Which stream of thought provides the best account for the phenomenon of moral motivation?" More specifically, I will be arguing in the negative as to whether or not anti-realist and hybrid views successfully avoid significant weaknesses of their own in attempts to develop plausible theories.

III. SIGNIFICANCE OF THE ISSUE

A. Significance of General Issues in Philosophy

Investigating the nature of moral properties has generally been the purview of armchair philosophy, however advances in neuroscience and psychology into understanding cognitive structures and emotions have opened new avenues of investigation. For example, if morals are mind-dependent and emerge from our emotions, then we should expect that people with abnormal emotional capabilities function in morally different ways than normal people. If empirical research confirms this link between brain structures and moral apprehension, then it would certainly qualify as strong support for an anti-realist moral understanding. Studying moral motivation is a promising area for progressing the metaethical dialogue. It provides valuable empirical research that could help philosophers answer questions about the nature of human moral experience.

B. Significance in Practical, Everyday Experience

The nature of morality is an issue of perennial importance. Human social groups are structured around codes of ethics which guide behaviour and govern expectations for praiseworthy and blameworthy behaviour. To both realists and anti-realists, it is important that moral values be objective in some sense. Societies need an agreed upon moral code to function properly. If moral realism is true, then these moral properties are mind-independent and people

can be held accountable for their moral actions. If non-realism is true, then moral properties are mind-dependant, and although it would certainly be possible for a society to create an epistemically objective set of moral values to follow, the guidelines for creating such a set of moral values are not exactly clear. Justification for holding someone accountable for their moral actions is also less clear on a non-realist view since mind-dependant ethics may vary across different cultures.

IV. PLAN OF RESEARCH

A. Methodology

Investigating the properties of moral facts requires a combination of armchair philosophy and evaluation of empirical research. Moral philosophy has had a long history of combining both abstract thought and empirical data because of the intrinsically practical nature of ethics in general. Empirical data, coming from people's common moral experience is what informs and drives abstract theorizing about the nature of moral properties. Most, if not all people are born with an innate ability to both understand and live out moral values in some form or other, and people's usage of moral facts in language points to a common-sense cognitive understanding of ethics in general. This experience of ethics is used as support for the metaphysical theory that moral facts are mind-independent and objectively true; the realist position. Metaphysical philosophizing evaluates whether a realist or non-realist position makes the most sense given the data.

In moral philosophy in general, and moral motivation specifically, the more empirical data philosophers have, the better. This paper will make use of empirical data regarding the nature and influence emotions have on moral motivation. If moral motivation is found to originate from mind-dependent factors such as emotion rather than an external, universal law,

such a discovery lends strength to an anti-realist position. Of course, finding strong empirical data to support a realist or anti-realist position does not settle the matter. Moral semantics are a crucial part of moral understanding, and a meta-ethical view will also have to accommodate how people talk about moral values. On moral realism, the meaning of "Lying is wrong" is semantically straightforward. "Lying is wrong" means that 'wrongness' is an objective, mind-independent quality, and 'lying' possesses this quality. Anti-realists have a less straightforward explanation of sentences like these, traditionally opting for explanations which revolve around "Lying is wrong" expressing emotional or rational dispositions towards 'lying'. Of course, anti-realists have more developed semantic theories than this, which we will explore.

The empirical focus of this paper will be moral motivation. Motivation is an excellent source to pull empirical data from because of its expected universality and neutrality with respect to specific moral codes. Generally, people both experience in themselves and expect from others, action when judging a moral 'good'. What people judge as 'good' can vary from person to person and culture to culture, but motivation to act on one's moral convictions is always expected. When one professes a moral judgment but fails to act, or acts in the opposite way, they are labeled as a deviant or a hypocrite. The desire and willingness to act on one's moral convictions is at the heart of ethical practice. Without it, questions about right or wrong simply become cognitive exercises both on realism and non-realism. If the nature of moral motivation has become clearer in the light of current research, it will certainly advance the field of metaethics.

I have chosen a few prominent exemplars of anti-realist views for this paper. Jesse Prinz has done important work regarding the emotional basis for action. He is a professor of philosophy well known for his work in emotion and moral psychology. He has written many

papers on the emotional nature of morality including “The Emotional Basis of Moral Judgments” (2006) where he argues that emotion is both necessary and sufficient for moral action.

Allan Gibbard’s “norm-expressivism” is another anti-realist position that will be examined closely in this paper. His work falls into the *hybrid* camp. Gibbard was one of the first of a new wave of expressivists who distance themselves from non-cognitivism. Norm-expressivism denies that moral judgments are reducible to psychological states and allows for moral sentences to express truth value. *Wise Choices Apt Feelings* (1990) is where Gibbard develops his theory, arguing that acceptance of rational ‘norms’ is what allows people to make normative judgments.

Lastly, I will be identifying a few philosophers who offer optimism for hybrid moral theories including Dorit Bar-on, and the purely cognitive account of Laura and François Schroeter.

B. Definitions

Moral Motivation. The desire to undertake a certain course of action (mentally or practically) based on one’s moral judgments.

Moral Realism. The position that ethical sentences express propositions that refer to objective features of the world independent of subjective opinion, considered to be true when they report these features accurately.

Moral Expressivism. Refers to a family of views according to which the meaning of ethical claims are to be understood in terms of whatever non-cognitive mental states those claims are supposed to express.[7]

Moral Sentimentalism. An anti-realist family of theories according to which, moral thoughts are either fundamentally sentimental (emotion-based), or that moral facts make essential reference to our sentimental responses, or that emotions are the primary source of moral knowledge.[8] [9]

Hybrid Theories: A meta-ethical position that combines one or more features from both realist and anti-realist theories.

Moral Judgment. A judgment that is (or expresses) a state of belief about right or wrong.

Cognitivism: A meta-ethical view according to which ethical sentences express propositions, and can therefore be true or false[10]

Non-Cognitivism: A meta-ethical view that ethical sentences do not express propositions, and cannot be either true or false. On a non-cognitivist view, moral knowledge is impossible.

---

[7] A.J. Ayer, C.L. Stevenson and R.M. Hare, are all major contributors to early non-cognitivist work. Allan Gibbard will be the main proponent of expressivist work discussed in this paper.

[8] Throughout this thesis, I will be referring to moral motivation, moral realism, moral sentimentalism, and moral expressivism as simply motivation, realism, sentimentalism, and expressivism unless being contrasted with the larger philosophical theories which bear the same names.

[9] Jesse Prinz is the main sentimentalist discussed in this paper.

[10] Cognitivism covers a broad range of moral theories, including moral realism, moral subjectivism, and error theory

C. Chapter Summary

In Chapter 1, my research will begin by interacting heavily with philosophers in the anti-realist tradition. As far as a realist position is concerned, I will give a brief overview, but for my purposes here, realism doesn't need to be explained in much more detail than I have already looked at in this introduction. The primary element of realist ethics under consideration here is that moral judgments express moral facts, and subsequently, that realism lacks a compelling solution to the phenomenon of moral motivation.

For the anti-realist position I will interact with *sentimentalism* primarily through the research of Jesse Prinz and Jonathan Haidt, both of whom have conducted empirical research into investigating the links between emotions and moral judgments, with a focus on moral motivation. Prinz, in his paper "The Emotional Basis of Moral Judgments" argues that there are significant links between moral judgments and the emotion centres of the brain, making emotion both necessary and sufficient for moral motivation.[11] Psychopaths are valuable to Prinz's research because they do not react to the same emotional stimuli as regular people, and if people who lack emotion are unmotivated to act on moral judgments, it might show that the link between moral judgments and motivation is emotional. Haidt argues similarly that his research identifies emotion as the basis for moral judgments and consequently, motivation.[12]

In Chapter 2, I will use Allan Gibbard's theory of normative judgment to explore the anti-realist expressivist position. Gibbard's theory covers normative judgments of every kind, and presents a unique solution for understanding moral motivation, arguing that our rational

[11] Jesse Prinz, "The Emotional Basis of Moral Judgments," *Philosophical Explorations*, (2006): 29-43.

[12] Jonathan Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgement," *Psychological Review* 108, no. 4 (October 2001): 814-834.

acceptance of a system of ‘norms’ provides grounds for motivation in that we are driven to act in accordance with ways which ‘make sense’ to us.

In Chapter 3, after interacting with both sentimentalism and hybrid expressivism, I will move into a more detailed critique of expressivism followed by examining cognitive alternatives to non-cognitive theories. By following a research plan as outlined, I hope to cover the development of non-realist ethical positions related specifically to moral motivation up to the present, which will give a clearer understanding about what research into moral motivation means for meta-ethics.

V. CONCLUSION

Moral judgments and moral action cannot be separated. Whether realist, anti-realist or hybrid, all agree that when a moral judgment is made, motivation to act on that judgment follows. Uncovering the nature and origin of this phenomenon will be invaluable to metaethical advancement, and will also help to shape people’s understanding and expectations of moral action from one another. This paper will explore and evaluate some of the best arguments anti-realists and hybrid theorists argue for, framing the metaethical debate in light of both current empirical and philosophical work.

Chapter One

SENTIMENTALIST APPROACHES TO MORAL MOTIVATION

I. OVERVIEW

A. Moral Realism

The claim "It is wrong to steal", like the claim "It is 25 degrees outside" appears to report a fact about the way the world is. People expect statements like these to report a true state of affairs, provided the world actually is as the claim dictates. The idea that moral claims report facts about the way the world is, and that at least some of these moral claims are true is what generally defines moral realism and moral cognitivism.

Semantically, people make moral claims in the same way that they make factual ones. "Dave's hair is brown" and "It is wrong to steal" both have the same sentence structure and are both generally understood to be making a claim about the *real* way the world is. The sentence "Dave's hair is brown" reports a natural property of the world (the brownness of hair) that exists, can be tested, and proven or disproven using the laws of science. Moral realists, on the other hand, have generally had difficulty explaining how moral facts fit into the natural world alongside scientific facts. While it is easy to come up with a tool that measures natural properties such as temperature, it is impossible to come up with a similar tool to measure amounts of "goodness" or "evil".

Of course, this difficulty might only go to show that despite semantic similarity, natural and moral facts are not identical in the same way. If moral claims do report facts about the natural world, but in a way that is undetectable by the natural sciences, what are the features of these moral claims that set them apart from natural facts? According to some philosophers, the

defining feature of moral claims is that they are essentially *motivating* in a way that natural claims are not. In other words, a moral claim is what it is, only if the person making the claim is motivated accordingly, or if such claims are expected to produce motivation in others. When someone makes the claim "Stealing is wrong" we hear the claim as a statement of fact, but we also expect that the person making the claim is also motivated not to steal (and intends to motivate others not to steal)–such that when they are presented with an option to steal, their motivation *not* to steal will prevent them from stealing. In addition, if I agree with the claim "It is wrong to steal", I should expect to feel motivated similarly.

In this chapter, I will be exploring the thesis that emotion is the basis for the motivation people feel to act on their moral judgments. This thesis is anti-realist and non-cognitive because it holds that moral judgments as well as the motivation to act on them come from the subject's emotional responses to situations most would consider part of the moral domain.

B. Emotion

Emotion is believed to be a strong candidate for explaining moral motivation by some psychologists and philosophers who have performed several psychological experiments designed to test the link between emotion and moral action. One study, conducted in the early 1970's by psychologist Stanley Milgram, instructed graduate students to request the seats of strangers on the New York subway. It is reported that almost all of Milgram's students did not even attempt this action because of how uncomfortable it made them feel. Milgram himself, not understanding what all the fuss was about decided to undertake the task himself. He reports his experience this way:

> The words seemed lodged in my trachea and would simply not emerge. Retreating, I berated myself: 'What kind of craven coward are you?'

> Finally after several unsuccessful tries, I went up to a passenger and choked out the request, 'Excuse me sir, may I have your seat?' A moment of stark anomic panic overcame me. But the man got right up and gave me the seat. A second blow was yet to come. Taking the man's seat, I was overwhelmed by the need to behave in a way that would justify my request. My head sank between my knees, and I could feel my face blanching. I was not role-playing. I actually felt as if I were going to perish.[13]

II. Prinz's Sentimentalism

Jesse Prinz states that recent work in cognitive science overwhelmingly supports a link between moral judgment-making and emotion. In his article "The Emotional Basis of Moral Judgments", he argues that emotion is both necessary and sufficient for a person to make a moral judgment. Prinz suggests that making a moral judgment is bound up with being motivated to act. He argues for a motivational internalist position[14], explaining that his view accounts for many of the features of internalism that internalists hold to be foundational.

Prinz's research begins with empirical data, which is then extended to philosophical reflection. Prinz believes that pure philosophical reflection is beneficial for approaching the conceptual question: "Can one possess a moral concept without having certain sentiments?", but recognizes that the concepts behind such questions are "graded, open textured, or polysemous, and philosophical intuitions are, correlatively divided."[15] Rather, Prinz prefers to start with a more simple question: "Do our ordinary moral concepts (the ones we deploy in token thoughts most frequently) have an emotional component?"[16] This gives the question a firm, empirical basis by determining if the emotion centres of the brain are active when a subject uses terms like

---

[13] Thomas Blass, *The Man Who Shocked the World: The Life and Legacy of Stanley Milgram* (New York: Basic Books, 2004), 174.

[15] Jesse Prinz, "The Emotional Basis of Moral Judgments," *Philosophical Explorations* 9, no. 1 (2006): 30.

[16] Prinz, "The Emotional Basis of Moral Judgments," 30.

'good' and 'bad' or 'right' and 'wrong'. Prinz follows by presenting 3 increasingly strong theses in support of the conclusion that there is an emotional component to ordinary moral judgments.

A. Thesis 1

Prinz's first thesis is that emotions and moral judgment[17] co-occur.[18] Starting with basic moral judgments, Prinz asks if there is an emotional component to everyday uses of the words 'good', 'bad', 'right', and 'wrong'. This thesis is regarded to be quite uncontroversial, as people often have a negative emotional response when they feel a moral norm has been violated. Supporting this claim are several studies which have explored this element of emotional psychology.

A 2003 journal article by Moll, de Oliveria-Souza, and Eslinger details a study where participant's brain activity was studied after the participant was asked to evaluate a sentence such as, "You should break the law when necessary" and then comparing the emotional response with more factual propositions like "Stones are made of water".[19] In both cases, participants only had to answer 'right' or 'wrong', but they found that emotional areas of the brain were active when they answered questions of a moral nature. Another study measured emotional brain states when participants were asked to split a monetary sum with another participant. Whenever the sum turned out unfavourably, negative emotional responses were observed.[20] Yet another study

[17] As mentioned earlier, Prinz believes that moral judgments are intrinsically motivating. His specific views on motivation will be explored later, but it is helpful to keep motivation in mind while discussing moral judgments in Prinz's work.

[18] Prinz, "The Emotional Basis of Moral Judgments," 30.

[19] Jorge Moll, Ricardo de Oliveira, and Paul J. Eslinger, "Morals and the human brain: a working model," *NeuroReport* 14, no. 3 (2003): 299.

[20] Alan G. Sanfey et al., "The Neural Basis of Economic Decision-Making in the Ultimatum Game," *Science* 300 no. 5626 (2003).

found emotional responses in subjects who were asked to look at pictures of politicians they disliked[21], and Prinz references even more studies which investigate this link between moral judgment and emotion.

These findings should be relatively unsurprising, as people generally expect a range of negative emotional responses when a moral norm or expectation is violated in basic moral situations. In addition to the kinds of moral situations the above studies explored, negative emotional responses are expected when people are treated rudely or with disrespect, and conversely, positive emotional responses are expected when giving someone a compliment or offering a gift. If these studies are accurate in their findings, then there is an empirical foundation to build on, linking emotion as a common and expected response to moral judgments. But does this link work the other way? Do emotions influence moral decision-making?

B. Thesis 2

Prinz's second thesis states that emotions influence moral judgments. A person who is already in a negative emotional state (angry, sad) will judge morally in a harsher way than they normally would when not under the effects of a negative emotion.

A 2008 paper details an experiment where subjects were given a series of descriptions, and are then asked to rate the wrongness of what was described. One example is as follows:

> Your plane has crashed in the Himalayas. The only survivors are yourself, another man, and a young boy. The three of you travel for days, battling extreme cold and wind. Your only chance at survival is to find your way to a small village on the other side of the mountain, several days away. The boy has a broken leg and cannot move very quickly. His chances of surviving the journey are essentially zero. Without food, you and the other

[21] Joshua Greene et al., "An fMRI investigation of emotional engagement in moral judgment" *Science* 293 no. 5537 (2001).

> man will probably die as well. The other man suggests that you sacrifice the boy and eat his remains over the next few days. How wrong is it to kill this boy so that you and the other man may survive your journey to safety?[22]

Researchers placed half of the participants at clean, orderly desks, while the other half were seated at filthy desks with half-eaten food on it. Participants sitting at the filthy desk were more likely to rate the hypothetical actions as wrong, compared with those sitting at the clean desk who did not judge the actions as wrong as often. These findings led the researchers to believe that negative emotions cause people to judge moral actions more harshly than when in a neutral or positive emotional state.

This experiment fails to consider another possibility. The questions asked of people in this study all lean towards the negative side of common decency without being outright wrong. Perhaps the negative emotions generated from sitting at the filthy desk simply drew participant's thought towards the mild 'wrongness' which was already present in the questionnaire. Finding out if emotion is sufficient to negatively influence moral judgments *when there is no reason to judge negatively* would be a more revealing study.

A 2005 paper by Thalia Wheatley and Jonathan Haidt investigates this. In this study, some participants were hypnotized to feel a flash of disgust whenever they heard the word 'often'. They were asked questions similar to those posed in Schnall's study, but with additional situations which included morally admirable characters. Wheatly and Haidt discovered that even in the situations that described morally admirable characters, hypnotized subjects were more likely to judge a moral wrong had occurred when experiencing feelings of disgust.[23] Prinz states

---

[22] Simone Schnall et al., "Disgust as embodied moral judgment," *Personality and Social Psychology Bulletin* 34, no. 8 (2008): 1108.

[23] Thalia Wheatley and Jonathan Haidt, "Hypnotic Disgust Makes Moral Judgments More Severe," *Psychological Science* 16, no. 10 (2005): 780-784.

that if these studies are to be believed, it shows that emotions are sufficient for a moral judgment, since being in a negative emotional state generates the belief that a moral wrong has occurred.[24]

Prinz claims that emotions are also necessary for moral judgments due to research on psychopaths. Psychopaths are often found to be deficient in certain negative emotions, particularly fear and sadness, but some argue that their emotional deficiencies encompass a wide range of both positive and negative emotions.[25] In addition to not experiencing emotional cues themselves, they also appear to have difficulty recognizing these same emotions in the facial expressions and speech of people they are with.[26]

There is an intuitive plausibility linking observed emotional distress cues with a motivation to act. Consider any charity that works with impoverished global communities. To motivate people to donate their time and money, these charities usually showcase images of sad-looking, emaciated people while beseeching the viewer to help, followed by happy and joyful scenes showcasing people who have been benefited by the charity. This strategy is successful in motivating people to give to the charity, and is undoubtedly more successful than not showing negative images. Psychopaths, being unable to recognize these emotional cues, would presumably feel unmotivated to assist the charity in the same way.[27] In an older study, Hervey

---

[24] Prinz, "The Emotional Basis of Moral Judgments," 31.

[25] A. Dawel et al., "Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy," *Neuroscience and Biobehavioral Reviews* 36, no. 10 (2012): 2288-2304.

[26] R.J. Blair et al., "A Selective Impairment in the Processing of Sad and Fearful Expressions in Children With Psychopathic Tendancies," *Journal of Abnormal Child Psychology* 29, no. 6 (2001): 491-498.

[27] This is an over-simplification of psychopathy that will be explored in more detail subsequently.

Cleckley described psychopaths as knowing that their actions are 'wrong' but not possessing an internal understanding of what 'wrong' actually means.[28]

> I think that psychopaths behave badly because they cannot make genuine moral judgments. They give lip-service to understanding morality, but there is good reason to think that they do not have moral concepts – or at least they do not have moral concepts that are like the ones that normal people possess. Psychopaths acknowledge that their criminal acts are 'wrong' but they do not understand the import of this word.[29]

Prinz's interpretation seems to be that the best explanation for the violent and unremorseful behaviour some psychopaths exhibit is because they do not have an internal experience of moral motivation to act rightly due to a lack of emotion, which would normally create a moral understanding. The relationship between emotions and psychopathy as just described is not an uncontested view, which will be explored when I move into a critique of Prinz's position, but for now, I will grant Prinz this understanding of psychopathy because it leads into the final thesis of his emotion-based theory.[30]

C. Thesis 3

Prinz final thesis argues that emotions are necessary for moral judgments in a synchronic sense. There would be a disconnect, he says, if a person were to say 'murder is wrong' and yet

---

[28] Hervey M. Cleckley, *The Mask of Sanity: An Attempt to Clarify Some Issues About the So Called Psychopathic Personality* (Augusta: C.V. Mosby Co., 1988).

[29] Prinz, "The Emotional Basis of Moral Judgments," 32.

[30] There are other theories which offer possible explanations of a psychopath's moral behavior. Lawrence Kohlberg argued that there are stages of moral development, with each successive stage broadening a person's moral perspective. With stages of moral development, a psychopath might only be able to achieve stage 4 (Ability to see abstract normative systems) but not make it to stage 5 (That moral contracts allow persons to mutually increase in welfare). Carol Gilligan criticizes Kohlberg's data, arguing that his research only utilized male participants, and fails to take female and non-gendered experiences of moral decision making into account. See Lawrence Kohlberg, "The Claim to Moral Adequacy of a Highest Stage of Moral Judgment," *Journal of Philosophy* 70, no. 18 (1973): 630-646., and Carol Gilligan, *In a Different Voice: Women's Conceptions of Self and Morality* (Cambridge: Harvard University Press, 1982).

feel no negative emotions when a murder happens.[31] Imagine a person who sees a video of ISIS militants beheading Christians in the Middle East. The person who watches that video might respond by saying something like, "Hmm, the action of killing I just witnessed is wrong because it is a non-universalizable maxim that reduces the fitness of the human race. ISIS shouldn't behead people because it goes against the human biological need to reproduce and multiply the species." Such a person affirms the proposition 'murder is wrong', but affirms it in a way that is confused or insincere according to Prinz. Conversely, a person who experiences a strong negative emotional reaction to murder without considering the facts of decreased utility that murder causes *can* be said to view murder as morally wrong.[32]

D. Summary

With these three theses, Prinz believes he has made a case for both the necessity and sufficiency of emotions for moral judgments. To recap, Prinz argues for sufficiency by referencing studies where participants are asked to make moral judgments while experiencing negative emotions (both from external and internal sources). These studies indicate that experiencing negative emotions are sufficient for causing a person to judge morality more harshly than when the negative emotions are not present. Prinz argues for necessity from the observations of psychopathic behaviour as well as common-sense. Psychopaths, who are said to be deficient in both experiencing emotions and observing emotional cues from others seem to not understand what morality is, which causes them to act in immoral ways without remorse. Arguing from common-sense, Prinz states that people who argue for or against moral actions for purely rational reasons cannot be said to truly understand concepts of 'rightness' and

[31] Prinz, "The Emotional Basis of Moral Judgments," 32.

[32] Prinz, "The Emotional Basis of Moral Judgments," 32.

'wrongness' while people who experience emotional reactions to the same events without considering the rational elements can be said to internally understand what 'rightness' and 'wrongness' are. With this combination of empirical data and common-sense understanding, Prinz feels that he has a good case to argue that emotions are both necessary and sufficient for moral judgments. To respond to Prinz's argument for the emotional basis of moral judgments, we will examine the case for the sufficiency and necessity of emotions for moral judgments separately.

III. Responses to Prinz

A. Sufficiency

Beginning first with the claim that emotions are sufficient for moral judgments. Prinz's statement on the matter is that empirical findings "suggest that we can form the belief that something is morally wrong by simply having a negative emotion directed towards it. *In this sense*, emotions are sufficient for moral appraisal"[33] The relationship Prinz suggests here appears to be causal in nature – in other words, external factors like the filthy desk and hypnotic suggestions were casually efficacious in bringing about the outcome of harsh moral judgments. Taken the way Prinz suggests, it does not seem that such evidence can be used to support a general understanding of how moral judgments normally form. A robust account of moral motivation must be able to account for a semantic understanding of moral judgments and the propositions which comprise them in ordinary circumstances. It is difficult to see how Prinz can argue that the harsh moral judgments made by the participants in the studies he references can be said to constitute genuine moral judgments when it is part of the experiment to arrange the

[33] Prinz, "The Emotional Basis of Moral Judgments," 31.

conditions to produce the desired outcome. To argue that the harsh moral pronouncements are genuine begs the question.

Karen Jones states this problem another way. Consider Samantha, a participant in the experiment who is told after the fact that her emotions had been subject to hypnotic suggestion before making the harsh moral judgments that she did. Now that she is aware of the hypnotic suggestion, the researchers ask Samantha if she would like to revise any of the moral judgments she had made. Samantha refuses, stating that if she felt disgust, that fact is a good enough reason to make the judgment she did, harsh or not.[34] There are two interpretations of this behaviour: Prinz is correct, and emotions are sufficient for a moral judgment, and we should expect Samantha to justify her moral judgments in this manner, or, a genuine moral judgment has not been made.

Contrast this example with another hypothetical case of Steve, an individual who makes a moral judgment based on an emotional reaction (without the use of hypnotic suggestion). Steve is confident that his moral judgment is correct, and that his emotional state is justified, but in response to some relevant reason, even if he is not cognisant of what that reason might be. There might be a problem with Steve's moral judgment, but the problem is one of competence rather that a conceptual one like in the case of Samantha. Hanno Sauer suggests that without linking reasons, one cannot properly articulate moral concepts.[35] Sauer's theory seems to align with the

[34] Karen Jones, "Metaethics and emotions research: A response to Prinz," *Philosophical Explorations* 9, no. 1 (2006): 48.

[35] Hanno Sauer, "Psychopaths and Filthy Desks: Are Emotions Necessary and Sufficient for Moral Judgment?," *Ethical Theory and Moral Practice* 15, no. 1 (2012): 95-115.

common-sense idea that Samantha is not producing a genuine moral judgment since she fails to be swayed by the relevant features of her situation.

B. Necessity

It is not far-fetched to expect that most people would doubt Samantha's competence as a moralizer. In Prinz's defense of the necessity of emotion for moral judgments, he asks if most people would consider someone to really understand what morality is if they felt no emotion tied to their moral judgments. The same criticism can be levelled against those who make moral judgments based purely on emotional reactions to a situation. How would such a person be able to enter into a discussion on moral issues when the sufficient cause for their judgments is how they feel at the time? Sauer, when speaking of incompetent moralizers says that such people *attempt* to make a moral judgment, but they cannot be regarded as succeeding because they judge in the wrong way.[36]

Clearly, making an authentic moral judgment requires rational consideration of the features relevant to the moral situation in question. Justification is relevant here, and while considering the relevant aspects of the moral case are certainly related to the justification of moral judgments, it should be equally clear that being upset at the filthiness of the desk one happens to be at is morally irrelevant. A moral theory must have the feature of differentiating between genuine moral judgments and those that only have the appearance of moral judgments. A moral theory that holds emotion as sufficient for moral judging lacks this feature.

Prinz himself seems to take moral reasons as being important when his argument turns to the necessity of emotions to moral judgments. In this section of his paper, he draws heavily on

[36] Sauer, "Psychopaths and Filthy Desks: Are Emotions Necessary and Sufficient for Moral Judgment?,".

psychopathic research to support his position. Although psychopaths utter sentences that sound moral, "Killing is wrong" has the same sort of meaning to a psychopath as "You shouldn't wear bright colours to a funeral". In other words, psychopaths view moral judgments and conventional ones similarly and have difficulty differentiating between the two. Prinz accounts for this difficulty by claiming that psychopaths are unable to experience emotional cues themselves, and cannot recognize them in others.

C. Psychopathy Unpacked

Firstly, it is important to recognize that psychiatric research is far from conclusive for the thesis that psychopathy necessarily involves an emotional deficiency. It is true that studies have been done which suggest that a psychopathic individual lacks responsiveness to situations where empathy is expected as well as responses to emotions like fear, even in children diagnosed with psychopathy.[37] [38] It is also relevant to point out that although earlier research focussed on negative emotional cues like fear, sadness, and empathy, recent fMRI research has been done that links deficiency in positive emotions like joy to psychopathy.[39]

i. Psychopaths Experience Emotion

Despite the volume of evidence supporting the claim that psychopaths don't experience emotion, the research is not conclusive. A recent study *instructed* psychopaths to empathize with

[37] Jean Decety and Laurie R. Skelly, "The Neural Underpinnings of the Experience of Empathy: Lessons for Psychopathy," in *The Oxford Handbook of Cognitive Neuroscience: Volume 2: The Cutting Edges*, ed. Kevin N. Ochsner and Stephen Kosslyn (New York: Oxford University Press, 2013), 228-243.

[38] R.J. Blair, "A cognitive developmental approach to mortality: investigating the psychopath," *Cognition* 57, no. 1 (1995): 1-29.

[39] Amy Dawel et al., "Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy," *Neuroscience & Biobehavioral Reviews* 36, no. 10 (2012): 2288-2304.

a victim when watching a video of a person being harmed.[40] The research team instructed the psychopaths to empathize with the harmed individual, and to imagine the physical sensations the harmed individual was feeling. Remarkably, the brain scan results showed that emotional centres of the brain related to empathy were activated when the psychopath was instructed to empathize, and the area of the brain associated with pain was activated when instructed to imagine how the victim physically felt. The conclusion of the research team was that psychopaths can activate emotional and empathetic responsiveness at will rather than the automatic response normal people have.

Despite this information, let us grant Prinz the claim that in the cases he references, the psychopaths did not feel the requisite emotions normal people would have. If granted, do these studies suggest what Prinz claims, that emotions are necessary for moral judgments? Eyal Aharoni, Walter Sinnott-Armstrong and Kent Kiehl suspected that the volume of research done on testing the moral understanding of psychopaths had been faulty.

Aharoni et al. cite early psychopathic research as the primary source for the view that psychopaths have an inability to process emotions correctly. In the mid 1990's R.J. Blair and his colleagues tested psychopaths using a well-known moral reasoning questionnaire known as the Moral-Conventional Transgressions task (MCT). This survey is meant to discover what a participant considers relevant for a moral transgression (e.g., pushing someone) to have occurred rather than a conventional one (e.g., spitting in public). Obtaining psychopathic participants from prisons and mental hospitals, Blair and colleagues discovered that psychopaths failed to

[40] Harma Meffert et al., "Reduced spontaneous but relatively normal deliberate vicarious representations in psychopathy," *Brain* 136, no. 8 (2013): 2550-2562, accessed August 24, 2015, http://dx.doi.org/10.1093/brain/awt190.

make a normative distinction between moral and conventional wrongs, while controls could make the distinction.[41] Aharoni feels that due to this this major study, subsequent researchers began adopting the view that psychopaths are deficient in moral reasoning abilities.

ii. Psychopath's Evaluation of Moral/Conventional Wrongs

The detailed results of Blair's study are fascinating. Blair found that psychopaths did not rate both kinds of wrong acts as permissible, but both as *morally wrong*. Blair's team interpreted these results as a manifestation of the psychopaths hoping to answer the questions 'correctly' in order to perhaps be released from incarceration early or obtain some kind of benefit or special treatment during their sentence. In other words, the psychopaths could not distinguish between moral and conventional wrongs, so they just marked everything as morally wrong, hoping that those answers would give them a favourable perception in the eyes of the researchers. Although this interpretation makes sense, it was never tested further.

Aharoni and his team decided to re-administer the MCT test in a different way in order to see if Blair's results could be duplicated.[42] To minimize the effect of having the participants think they needed to answer according to the researcher's wishes, Aharoni's group told them that exactly half of the 'wrongs' listed were moral, and the other half conventional. If Blair's results were accurate, psychopaths should answer randomly and non-uniformly with other participants.

---

[41] R.J. Blair et al., "Is the psychopath 'morally insane'?," *Personality and Individual Differences* 19, no. 5 (1995): 741-752.

[42] Eyal Aharoni, Walter Sinnott-Armstrong, and Kent A. Kiehl, "Can Psychopathic Offenders Discern Moral Wrongs? A New Look at the Moral/Conventional Distinction," *Journal of Abnormal Psychology* 121, no. 2 (2012): 484-497.

Aharoni's team found that psychopathic participants accurately classified 82.6% of the 'moral' wrongs correctly as opposed to 92.5% for control groups. With consideration from other factors being tested in this study, the researchers concluded that "psychopathic participants performed no worse than controls in moral reasoning tasks" and also that psychopaths accurately rated moral wrongs with much higher severity than non-moral unpleasant wrongs.[43]

Unlike the Meffert experiment, Aharoni's team did not coerce an emotional response from the psychopathic participants. Because Meffert had to instruct her participants to become emotionally active before the emotion centres of their brains became active, it is reasonable to assume that Aharoni's participants were not emotionally involved, yet they accurately identified moral and conventional wrongs. The ability of psychopaths to accurately identify and distinguish moral wrongs from conventional ones contradicts Prinz's claim that emotions are necessary for making a moral judgment.

iii. Prinz's Possible Responses

Prinz's response to these findings would likely be to repeat his earlier stance: a critique of the sincerity of a person who makes a moral judgment and yet feels no emotional response to the judgment. I find this hard concept to universalize. Granted, there will be cases where a subject gives lip service to moral judgments, verbalizing the rightness or wrongness of an action while internally not caring one way or the other, while simultaneously feeling no emotional cues. Such a person is rightly ascribed insincerity as a moral judger in cases so described. But does it follow that *any* person who lacks emotional cues while making moral judgments demonstrates insincerity?

[43] Ibid.

Imagine two people considering racism. They understand that in many places in the world, people use derogatory terms to degrade people of different races, or as a tool to insult. Person A has experienced racism and the negative feelings that go along with being the victim of racially charged hatred. Person A judges that racism is wrong, and is motivated to act against racism due in part to their emotional history with the subject. Person B has had a sheltered life and has not been the victim of racism personally, although she knows it exists. When Person B considers racism, she also judges that it is wrong, primarily because it makes sense to her not to judge a person's worth and ability on the colour of their skin, and she understands that a person's worth is not bound up in these things.

Both Person A and Person B judge that racism is wrong, one drawing from their emotional experiences with racism, and the other from their understanding of a person's inherent value. The question is whether Person B less competent as a moralizer than Person A. It seems to me that Person B is just as competent a moralizer as Person A and no less resolute in taking a stance against racism.

Prinz argues that emotions are both necessary and sufficient for moral judgments in addition to providing the motivational force a subject experiences to act on their judgments. As we saw, the necessity claim is called into question when we look at some of the further work done with psychopathic offenders which shows that they can accurately differentiate between moral and conventional wrongs. Further, the sufficiency claim fails to hold up when we consider the competence of a moralizer who rejects the relevant circumstances and only makes her judgments based on her emotional state of mind at the time.

At best, the conclusions we can draw from Prinz's work alongside the empirical data drawn from other sources indicate that human beings often experience emotional cues alongside

their moral judgments, and that these emotional cues can influence a person's motivational drive. Based on the material explored above, it would be a mistake to draw a conclusion that definitively places emotions in a position of necessity or sufficiency for moral motivation.

Chapter 2

ALLAN GIBBARD'S NORM-EXPRESSIVISM

In the first chapter, we considered the anti-realist claim that moral motivations arise from emotional sentiments brought on by the consideration of moral judgments. In this chapter, we will look at an alternative anti-realist view on moral motivation from the work of Allan Gibbard. In his book *Wise Choices, Apt Feelings*, Gibbard lays out his theory of normative judgment, where he attempts to show that our common-sense view of rationality explains why we have motivation to act on our moral judgments.

I. Norm-Expressivism

A. Normative Judgment as Rationality

To understand Gibbard's theory we must begin with how he sees normative judgment used in general terms of 'rationality' before moving on to how he applies it to moral motivation. Gibbard's proposal starts with a consideration of morality. He defines morality broadly as the question of how to live, and narrowly as concerning moral sentiments like guilt, resentment, and others. In the narrow sense, he agrees with Prinz that moral sentiments are necessary to morality, but disagrees with the sufficiency claim, saying that "To feel guilt or resentment is not in itself to make a moral judgment. A person can feel guilty and yet think he has done no wrong."[44] Due to this view, Gibbard diverges from Prinz's sufficiency claim regarding emotional sentiments to say that it is not the sentiments that make a moral judgment, but judgments regarding what moral sentiments it is rational for a person to have instead. To make this claim, we need to understand what Gibbard means by "rationality". Gibbard begins his

[44] Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge: Harvard University Press, 1990), 6.

work by asking what it means to call something rational vs. irrational, and after laying the groundwork of that question, to move to morality.

> To call a thing rational is to endorse it in some way. That suggests a scheme for getting at the meaning of the term. Instead of trying to define a property "rationality" by giving conditions under which a thing would have that property or lack it, start with the use of the term. Fix on the dictum "To call a thing rational is to endorse it," and search for a sense of 'endorse' for which the dictum holds true.[45]

Gibbard wants to use the word "rational" in the way most people do when they express their understanding and affirmation for something. It is the way people think when they say "that makes sense" or what we want when searching for the 'best course of action'. When applied to feelings, this sense of rationality encompasses viewing anger as warranted in certain situations and not in others. For an act to be rational, it must be the thing that makes the most sense to do given the situation, or at the very least it must be a course of action that makes enough sense to not be considered ill-advised. Gibbard tightens this broad definition somewhat to say that an act or way of feeling is rational because one accepts norms that permit it. Described this way, Gibbard is not analyzing what it means for something to *be* rational, but what it means for someone to *judge* that something is rational. The analysis Gibbard presents is therefore non-cognitive in a technical sense in that (on Gibbard's theory) to call a feeling or action rational is not to state a matter of fact about the nature of the thing; but as we will see later, factual language is not abandoned, and is important for Gibbard's norm-expressivist theory.

[45] Gibbard, "Wise Choices, Apt Feelings," 6.

i. A Naturalistic Framework for Norms

Gibbard recognizes the challenges that face those who try to reduce normative talk into a naturalistic framework. When done, reasons and meanings start to look unrecognizable from how people experience them. The challenge is to create one coherent picture for 3 kinds of facts:

1 Naturalistic facts: us as a part of nature, our acts and thoughts and feelings as they might be understood in a natural science

2 Normative facts: what it is rational to do and think and to feel, what it *makes sense* to do or think or feel

3 Facts of meaning: what our words mean in general, and particularly, what our normative claims mean – what it means to say it is rational to do or think or feel such-and-such[46]

On Gibbard's analysis, all strict facts will be naturalistic, facts of meaning will also turn out to be real facts, and so will also be naturalistic, and normative facts are not really facts at all, although there will be facts about what people do when they make normative judgments. Because the analysis rests on a naturalistic account, Gibbard starts with a normative psychology to explore how this naturalistic picture works. The key for Gibbard's normative psychology is that human beings live socially, and that our minds are designed for social life. Moral nature enters the picture when we consider that in order to live socially with one-another, a system of *coordination* must be in place to facilitate our actions.

> Primitive human life is intensely social. In the conditions under which we evolved, anyone's prospects for survival and reproduction depended crucially on the beneficial human bonds he could cultivate. Human cooperation, and coordination more broadly, has always rested on a refined network of kinds of human rapport, supported by emotion and thought. A person sustains and develops this network, draws advantages from it, and

[46] Gibbard, "Wise Choices, Apt Feelings," 23.

> on occasion keeps his distance from aspects of it. He does these things only in virtue of a refined configuration of emotional and cognitive dispositions. It is this picture of human biology that might represent us, however crudely, still as recognizably human.[47]

ii. Psychology of Rationality

We turn to the psychology of what it means to appraise something as rational. Start with a simple example of a woman who has evidence that her husband has been cheating on her. The question of whether it is rational to believe he is truly faithful depends solely on her evidence, and nothing else. Considerations of how she might be emotionally hurt if his infidelity is true, or how it strengthens their relationship if she thinks him faithful, we think, should not inform the rationality of her believing him unfaithful or not. This shows that the rationality of a belief and the desirability of a belief are not the same thing. Similarly, it would not be a good idea for one of Cleopatra's servants to express their anger at her, even if she slighted him and thus "made sense" for him to be angry. Perhaps he wants to rise in her favour and recognizes that expressing his anger would hinder this goal, or perhaps cause him physical harm.[48] We see this same thinking at work through the phrase "Don't shoot the messenger". In cases where this applies, it "makes sense" for the recipient of bad news to be angry, and yet not make sense to hurt the one who brings the bad news.

When understanding the word 'rational' in this sense, we can understand a little more clearly the competing uses of rationality. In the case of the servant and Cleopatra, we see that it is rational for him to be angry at Cleopatra for the slight, but that it is also rational for him to want to restrain his anger for the purposes of rising in her good graces. Thus, we see that it is

---

[47] Gibbard, "Wise Choices, Apt Feelings," 26-27.

[48] Gibbard, "Wise Choices, Apt Feelings," 37.

rational for him to be angry, but also rational for him to *not want to be angry*. A thing being rational to believe is different from *wanting* to believe it.

B. Emotions and Rationality

Talk of beliefs being rational or irrational seems reasonable, but talk of emotions being rational or irrational as we have seems suspect. True, people identify emotions as rational or irrational in common language as, "it makes sense for her to be angry" or "it doesn't make sense for her to be angry", but on closer examination, people may want to exonerate the charge of irrationality because (we think) emotions are not under voluntary control in the same way that reasoned thinking is. A person might be able to suppress their emotional response, but cannot conjure up any emotion at will, as if emotions could be controlled completely.

i. Voluntary and Involuntary Emotions

Gibbard agrees that emotions cannot be summoned and experienced at will, but denies that "Only the voluntary can be appraised as rational or irrational."[49] He identifies beliefs as prime examples of what he means. We can appraise beliefs (like a belief that the Earth is flat) as rational or irrational, even though people cannot simply choose to genuinely believe something at will. Where then does the key to rationality lie? Take actions: I choose to not procrastinate on my work, and get it done in a reasonable amount of time. This is an action that most, if not all would consider rational. It is not rational simply because I voluntarily chose to undertake it, instead, it is rational because of the intention and preferences that stand behind the action. I get my work done early and in a reasonable amount of time because I have a desire to not rush my

[49] Gibbard, "Wise Choices, Apt Feelings," 38.

work at the last minute and do a bad job, and it is based on this intention and others surrounding my action that it is appraised as rational.

Although Gibbard has more to say on rationality, what we have looked at gives us a good enough understanding to move on to how rationality and ethics are connected on his view. Ethics of course has different views on what right living actually is. Kantian approaches see morality as living in the purest form of practical rationality. If accepted, it wouldn't make sense to ask, "Is doing the right thing always rational?" because "the right thing" is always "the rational thing". On more narrow conceptions of morality however, doing the irrational is not necessarily immoral. A person who is afraid to go out on a boat in the ocean because they fear falling off the edge of the world is irrational, but their decision to not get on the boat isn't immoral.

John Stuart Mill, using this narrow view of morality describes the nuances of moral judgement this way, "Morality pertains to what is wrong or not wrong, and to say that an act is wrong is to say that there ought to be a sanction against it, a sanction of law, of public opinion, or of conscience." Of course, Mill thinks that utilitarian considerations should judge moral acts, but we can appropriate his line of reasoning to the theory of rationality that Gibbard proposes.

Imagine that when Mill says that there ought to be a sanction on moral actions, we think of it as a sanction of rationality, not law. Perhaps not saying "thank you" to someone who has given you a gift is not something that we think requires legal action, but we do think they should feel guilty for not doing so. This leaves in Mill's claim that sanctions of public opinion matter, and thus allows us to formulate a definition of *morally wrong* being an act that a person has rational reason to feel guilty about, and one that makes sense for others to be angry at him for as well. I use the word 'angry' here not for the purposes of claiming that the proper emotional

response to moral wrongs is “anger”, but as a blanket term to cover feelings like, hurt, resentment, indignation, frustration, blame, etc.

Even this conception of moral wrongs needs further refinement however. There are two senses that an act can be wrong: objectively and subjectively. The difference is shown through examples like this: I am an honest banker and I do my due diligence to make sure that money laundering doesn’t happen with my accounts through all the checks and systems the bank sets up for me to use. However, despite my best measures, money laundering happens on one of my accounts. Thus, while allowing the deposit to go through turned out to be wrong in the objective sense, it was not in the subjective sense, since I had every reason to think the deposit was valid, due to my diligence. Therefore, an act is morally wrong objectively when all the surrounding facts both knowable and unknowable are taken into account, but wrong in the subjective sense only if it makes sense to consider the subject culpable in light of their particular knowledge and intention. Obviously then, Gibbard’s analysis works well for acts that are morally wrong in the subjective sense, but not in the objective sense. After all, it would make no sense to hold the banker morally responsible for a fraudulent deposit that slipped through the cracks despite his best efforts to watch for and stop fraudulent transactions.

Having a moral theory that works in this subjective sense is certainly advantageous: it allows for moral guidance even when ignorant of the relevant facts, as one can use the facts they *do* know to pass the moral judgment that makes sense. Having a moral theory that covers the objective sense is unlikely. In most (if not all) moral situations, we are ignorant of at least some relevant facts that bear on ethics. We might think that if we were omniscient, with all knowledge of the relevant facts, an objective theory of ethics would be used, but if this were so, and we really did know all the relevant facts, an objective theory would become redundant since such a

person would already have the criteria for rightness in the subjective sense. Thus, when Gibbard uses the terms 'right' and 'wrong', he does so in the subjective sense.

C. Summary of the Theory

This all comes together under what Gibbard calls 'norm-expressivism'. First, a discussion about what norm-expressivism means in general, and then how it applies to moral motivation.

To call something rational, or believe that it is rational is to accept norms that permit it.[50] By 'norm' Gibbard means some kind of rule expressible by an imperative. The following story expresses how this works:

> Delilah, suppose, is pondering whether various of Samson's acts, beliefs, and feelings are rational. What is it for her to come to an opinion? It is to come to accept norms. When Samson destroys the Philistine temple, Delilah considers that rational if and only if she accepts norms that permit, for Samson's situation as she takes to be, destroying the temple. Perhaps she accepts the norm "When in the hands of one's enemies with no hope of escape, kill as many of them as possible, even if you must kill yourself in the process." Then if she believes that Samson is in the hands of his enemies with no hope of escape, and that destroying the temple will kill as many as possible, she thinks his action rational. Earlier, Samson believed Delilah loyal. Delilah thinks this belief to have been rational if and only if she accepts the right sorts of norms for belief. She must accept norms that, for Samson's situation at the time as she now conceives it, permit believing one's woman loyal. Samson hates the Philistines, and Delilah considers his hatred rational if and only if she accepts norms that, for his situation, permit such hatred.[51]

Note that throughout the story that Gibbard tells here, he doesn't comment on whether Samson *is* rational. Rather, his theory is about what it is to *consider* one rational. An observer considers the beliefs or actions of another rational if the observer accepts norms that permit the belief or action given the circumstances.

---

[50] Gibbard, "Wise Choices, Apt Feelings," 46.
[51] Gibbard, "Wise Choices, Apt Feelings," 46.

i. Relation to Ethics

Let us turn back to morality now. Recall that Gibbard is using the narrow moral terms of "blameworthy" and "wrong". If combined with the notion of rationality, we have a theory that assigns wrongdoing to an individual when it is rational for an observer to think the agent should feel guilt over an act and for others to resent that agent.

Now, considering what it means to be morally wrong in the 'subjective sense'. Here, Gibbard's standards dictate that an act an agent does is in the moral wrong, and generate feelings of guilt and resentment, if the agent chooses not to rule out alternatives that would prevent those feelings. Therefore, an agent thinks that an act is wrong when they accept norms for guilt and resentment that would allow others to think of them as deserving of guilt and resentment if they were to act in that way.

We can now see how the expressivist views ethical thought broadly, and how this informs motivational dispositions towards ethical acts. With non-moral acts people are motivated to act in accordance with norms that they hold to be rational. Garrison wants to live at a certain level of material comfort, so he searches for a job that pays in the range that would allow this sort of lifestyle. Aaron wants to get into better shape, so he chooses to eat less unhealthy foods, and to exercise regularly. The details of how the result is pursued varies depending on the particular set of norms each holds; Garrison might only apply to certain types of jobs, and Aaron may exclude certain foods and forms of exercise from his routines, but at the foundational level, motivation to act is not due to the norms themselves, but a desire for the end result.

On expressivism then, if we are motivated to act morally, it is because we have moral interests and concerns that we care about. Ethical action seems to be motivated by many possible concerns: Justice, benevolence, concern for other's well-being, respect, devotion to God etc. Perhaps these concerns boil down to one overarching concept; Jesus taught, "Love the Lord your God with all your heart and with all your soul and with all your mind."[52], and concluded that the entire Law and teaching of the Prophets was based on this command, and the command to love one's neighbour as oneself. Love may be the root of ethical concern, but it is uncommon that people filter their ethical motivations through 'love' consistently and in all situations. Indeed, it seems that different moral concerns grip people strongly or weakly that vary from person to person.

D. Review

Gibbard's expressivist theory analyzes judgments in terms of rationality based on how they express an individual's non-cognitive attitude of an acceptance of 'norms'. Further, Gibbard claims that *all* normative judgments are judgments such that acceptance or rejection of norms predicates them, meaning that all moral judgments (being normative) are covered by his analysis. An individual calls an action rational when they accept a system of norms which allows it, and irrational when one accepts a system of norms which forbids it.[53]

Gibbard's analysis covers moral judgments, also in terms of rationality. Actions are wrong when it meets standards (either intentionally or by negligence), such that a person in a normal state of mind would find the agent blameworthy. Standards for finding an agent

[52] Matt. 22:37 NIV.

[53] This is a rough approximation of Gibbard's full theory, suitable for the purposes of this paper. Gibbard allows for situations where a speaker may not have a system of norms in mind when making a judgment, and also deals with more complex judgments where sets of 'norm-world' pairs rule each-other out to arrive at a state of mind that suits the judgment in question.

blameworthy would be when it would be rational for the agent to feel guilty, and observers to resent the action of the agent. Because the standards for guilt and resentment are non-cognitive in nature, Gibbard's moral theory is non-cognitive with respect to moral judgments themselves.

Due to its reliance on the sentimental states of anger, guilt, and resentment to judge morally, Gibbard's analysis falls into the sentimentalist camp, as does Prinz, who we looked at in the previous chapter. What makes Gibbard's analysis of moral judgments different from Prinz's for our purpose of evaluating its efficacy for motivating moral action? Prinz's analysis rested solely on sentimental states to provide a motivating force for moral action through both the sufficiency and necessity of emotion to moral judgments. Such a theory holds that a moral judgment is purely an expression of the speaker's feelings towards the matter in question since sentiments do not have prescriptive truth values; falling into a non-cognitive, but subjectivist camp. Gibbard's analysis (and those of other non-cognitivists like him) also believe that norms lack truth value, but that a logic of norms is possible; thus falling into a non-cognitive, objectivist camp.

II. Evaluation of Norm-Expressivism

A. R.M. Hare and the Logic of Normative Sentences

Per Hare, normative sentences have three components: prescriptively, universalizability and supervenience.[54] Moral statements are prescriptions used to inform an action or sentence in response to the question, "What should I do?"[55] These are different from descriptive sentences – which say something about the way the world *is*, and imperative sentences – which tell the

[54] R.M. Hare, *Essays in ethical theory* (Oxford: Oxford University Press, 1989).

[55] R.M. Hare, *The language of morals* (Oxford: Clarendon Press, 1952).

hearer to *make* something the case. Moral sentences are a form of imperative then, as their goal is to persuade the hearer to a certain course of moral action. "Persuasions imply a lack of rationality to moral theories; therefore, using persuasion does not mean rationally replying to the question "What shall I do?", but rather it is an attempt to answer the question in a particular way."[56]

Moral sentences share universalizability with descriptive sentences per Hare. He states that when moral sentences use language like "ought" and "must", it is similar to when descriptive sentences use the words "all". In effect, the normative language of moral sentences must be compared to the logical operators "all", "some" or "It is necessary that" rather than being compared to predicates.[57] Since these rules govern moral sentences, they are universalizable.

Finally, supervenience is another feature that moral sentences share with descriptive ones. In ethics, this feature makes itself known through the question is about how moral properties relate to natural ones. Alternatively, one might claim of a certain moral feature like "love", that an act is loving *in virtue of* some requisite set of criteria. Thus, the "lovingness" of the action in this case supervenes on the relevant criteria.

Because normative sentences share these logical connectives with descriptive sentences, we can create and understand the logic of normative sentences in the same way we do descriptive

[56] Antonio Marturano, "Non-Cognitivism in Ethics," *Internet Encyclopedia of Philosophy*, http://www.iep.utm.edu/non-cogn/#H3.

[57] R.M. Hare, *Freedom and Reason* (Oxford: Clarendon Press, 1963).

ones. Hare states "that any formula of formal logic which is capable of an indicative interpretation is capable also of an imperative one."[58]

i. Relation to Norm-Expressivism

Gibbard, building on this foundation accepts the logic of norms while keeping an emotivist bent for the evaluation of moral judgments. One of the driving forces behind Gibbard's non-cognitivist view is that cognitivist evaluations fail to recognize that judging something as rational (or good in the case of moral judgments) must carry an endorsement or persuasive element on the part of the speaker. Feelings, he says, can be apt or not, and judgments of moral wrongs are when feelings of guilt and resentment are apt. Norms invariably assist with social cooperation and so while descriptive sentences deal with representations of the world, normative sentences make social cooperation stable, and keep them from being strongly guided by other societal or environmental factors. The key element here is that when one accepts a "system of norms", it justifies a psychological theory of meaning. Norms are an ambiguous state of mind, even to Gibbard, so they are not fully understandable. The result is a theory that embraces some uncertainty: on the one hand, moral judgments lack truth value, but they express the objective state of mind of the speaker.

B. The Frege-Geach Problem

Any expressivist moral theory must provide a meaningful response to the Frege-Geach problem (also known as the "embedding problem"). In brief, the Frege-Geach problem is that semantically, moral statements in complex sentences create problems for the expressivist to explain. For example, the sentence "Bullying is wrong" has the same meaning, regardless of if it

[58] R.M. Hare, "Imperative Sentences," *Mind* 58, no. 229 (1949): 21-39.

is a sentence on its own, or as the antecedent of a more complex sentence such as, "If bullying is wrong, then teaching your brother how to bully is also wrong." This must be true because one can derive "teaching your brother how to bully is also wrong" from both sentences using *modus ponens*. Yet despite this, "Bullying is wrong" does not express anything when it forms the antecedent of the conditional (in a relevant sense) and this issue does not find an answer in expressivist analysis.

i. Gibbard's Response to the Frege-Geach Problem

Gibbard's attempt to answer the Frege-Geach problem involves a variant of possible world semantics that he calls "factual-normative worlds." Factual-normative worlds are an ordered pair where "w" is a possible world (or set of facts) and "n" is a complete system of general norms. The pair constitutes a creedal-normative state completely opinionated.[59]

> According to Gibbard, any particular normative judgment holds or not, as a matter of logic, in the factual-normative world. That is, the pair is a set of sound and complete norms where, for each possible human behaviour, we can state the normative status (Forbidden, Obligatory, or Indifferent) associated with it. In this way each individual can understand the normative qualification of his or her action.[60]

Returning to the previous example in *modus ponens* form:

1. If bullying is wrong, then teaching your brother how to bully is also wrong.
2. Bullying is wrong

Therefore, teaching your brother how to bully is wrong.

[59] Allan Gibbard, "Wise Choices, Apt Feelings," 95.

[60] Antonio Marturano, "Non-Cognitivism in Ethics," http://www.iep.utm.edu/non-cogn/#SH4b.

Imagine an observer who is unsure both factually and normatively. This observer will evaluate the rightness of the action by ruling out possible actions that conflict with the facts of the syllogism and any norms that conflict with the normative judgment being valid.

Applying this to our syllogism, we can see that the first premise rules out all combinations where it is not wrong to teach your brother how to bully. The second premise rules out norms and facts where it is not wrong to bully. Thus, both premises together rule out any facts and norms where it would not be wrong for one to teach their brother how to bully.

What it means then for a sentence to be valid in a factual-normative world then means that for a sentence containing a normative predicate, there is a n-corresponding description that makes the word "rational" apply to that norm. Therefore, Gibbard argues that for any sentence containing an embedded normative predicate, there exists a descriptive sentence S_n which is constructed by replacing the normative predicates in the original S. Then, we can say that S holds iff S_n holds in a possible world. Through this method, Gibbard bypasses the Frege-Geach problem rather than providing a direct answer.

C. Problems for Norm-Expressivism

As pointed out previously, norm-expressivism is very clearly an emotivist theory even though it includes and appropriates methods from other fields such as biology, psychology, and decision theory that turn it into quite a different thing than Prinz's emotivist theory that we looked at in the previous chapter. As we saw with Prinz's theory on moral motivation, pure emotivism has several serious problems. Gibbard is aware of these issues, and formulates his theory to avoid and amend some of the more serious difficulties as we saw with his approach to the Frege-Geach problem. In what follows, I will explore some of the new difficulties Gibbard's norm-expressivism creates that are not adequately addressed in his method.

i. Is Norm-Expressivism an Expressivist Theory?

The first problem that Gibbard runs into is that some of his cleverer methods of avoiding common expressivist pitfalls risk turning his theory into a *non-expressivist* theory, despite his best intentions. Expressivism is a family of views according to which the meanings of claims are to be understood in terms of whatever non-cognitive mental states those claims are supposed to express.[61]

When Gibbard circumvents the Frege-Geach problem, he asserts that holding a moral normative belief is to be in a particular state of mind, and uttering a normative moral statement is to express that state of mind. When one does so, they rule out any possible combinations of normative judgments that would conflict with the original normative moral judgment. Let's re-examine the modus ponens argument we created in the previous section.

1. It is wrong to bully.
2. If it is wrong to bully, then it is wrong to teach your brother how to bully.
3. Therefore, it is wrong to bully.

Traditional expressivist responses have trouble with arguments like this. Although the first premise could be interpreted as expressing an attitude or state of mind towards bullying, the second premise does not seem to be expressing any sort of attitude towards bullying. In the first premise, the use of the word "wrong" expresses an attitude, but the use of "wrong" in the second premise uses it with a different meaning, therefore making the argument commit the fallacy of equivocation, with expressivists being unable to explain how the argument is valid. As we saw,

[61] James Sias, "Ethical Expressivism," *Internet Encyclopedia of Philosophy*, http://www.iep.utm.edu/eth-expr/.

Gibbard's response to this problem is to say that the first premise rules out any combination of facts and norms where bullying is not wrong. The second premise rules out the intersection of combinations of facts and norms in which bullying is wrong with the set of combinations where it is not wrong to teach your brother how to bully. Together, the premises rule out all the combinations of facts and norms where it is not wrong to teach your brother how to bully, including any combination that the conclusion rules out. This goes to show how the modus ponens can be valid.

This is a clever way of dealing with the issue, but it quickly becomes clear that Gibbard must leave the expressivist camp in order to make the explanation work. Walter Sinnott-Armstrong compares Gibbard's solution to a similar one in the realm of causation.

> Suppose someone claims that to believe (or to assert) that *a* causes *b* is just to be in (or to express) a state of mind which consists in ruling out various combinations of causal laws with non-causal states.[62] The validity of modus ponens with causal judgments would then be explained just as Gibbard explained the validity of modus ponens with normative judgments. However, this analysis and explanation are compatible with a fully realistic view of causal claims as claims about an independent world. Similarly, Gibbard's analysis of the content of normative judgments and his explanation of the validity of modus ponens seem just as fully compatible with a fully realistic view of normative claims as claims about an independent world.[63]

This comparison doesn't show that Gibbard's explanation is wrong, but it does illuminate the fact that his method deviates from being purely expressivist. Gibbard does consistently hold to the expressivist claims that normative moral judgments have truth values, and he explains that

---

[62] The theory being referenced here is from Davidson in "Causal Relations", *Essays on Actions and Events* (New York; Oxford, 1980). Davidson claims that to say "a causes b" is to claim that there is some causal law linking a and b without committing to any particular causal law, so Gibbard claims to say "a is rational" is to express acceptance of some norm that permits it without committing to any particular norm.

[63] Walter Sinnott-Armstrong, "Some problems for Gibbard's norm-expressivism," *Philosophical Studies* 69, no. 2 (1993): 297-313.

motivation to act on these moral claims is to accept a norm or set of norms. These expressivist claims create major problems of their own.

ii. Normative Governance

More relevant to our investigation of moral motivation is the issue of normative governance. Gibbard's norm-expressivism reports that when one accepts a norm, they are then motivated or disposed to act in accordance with the norm. Gibbard explains normative beliefs in terms of fact-norm combinations that rule out certain actions. But he also must show how a belief can be generated in a subject's mind based on the fact-norm combinations. It is easy to fall into a simple form of internalism to explain this link: A person who accepts a normative judgment is a person who must be motivated by it.

It is quickly obvious that this simple internalism cannot be true. I can possess the belief that I am (morally and rationally) permitted to drive in excess of speeds of 200 km/hr on the Autobahn (a normative judgment of permission), and yet not feel motivated to do so. If I have some other overriding reason to do so (I have an injured friend that has to get to the hospital as quickly as possible) then I would drive in excess of 200 km/hr, but this only shows that simple internalism holds for overriding reasons. When it comes to ethics, I can believe that giving to charity is the right thing to do, but if I am convinced that my small donation wouldn't make a large difference, and this causes me not to feel guilty, then I might not be motivated to do so. Gibbard claims to reject the sort of simple internalism that gives rise to counterexamples like these.

Accepting a norm is whatever psychic state, if any, gives rise to this syndrome of avowal and governance by it. This governance is described by Gibbard as: "Working out in community

what to do, what to think, and how to feel in absent situations, if it has these biological functions, must presumably influence what we do, think, and feel when faced with like situations. I shall call this influence *normative governance*." [64]

Instead of an internalist motivation, Gibbard claims that his method shows that motivation comes from community-based, normative discussion. When someone does something out of habit, or from a physiological need (like taking an afternoon nap), the motivation to do so has not arisen out of normative discussion, and so it is not necessary for them to accept norms that permit taking an afternoon nap. By contrast, if we discuss the prospect of lending aid to Syrian refugees and decide that we should help; when this motivates action, then we are normatively governed.

This reliance on normative governance doesn't seem to free Gibbard from the simple internalism he claims to reject; that acceptance of norms implies motivation. The kind of relevant motivation is more specific under Gibbard's analysis, but Sinnott-Armstrong points out that this only adds the incredulity of thinking that this narrow sort of motivation must accompany *any* sort of norm acceptance. Again, I can believe that giving to charity is the morally right thing to do and yet feel little desire to act on it. Discussion about the matter may determine my position on the matter, but it's hard to see how that is an expressivist trait.

Gibbard claims that norm acceptance requires both governance *and* avowal, so if there is a problem with governance, the whole method is suspect; but there are also difficulties with norm avowal. "To accept a norm, we might say, is in part to be disposed to avow it in

[64] Allan Gibbard, "Wise Choices, Apt Feelings," 75.

unconstrained normative discussion, as a result of the workings of demands for consistency in the positions one takes in normative discussion." [65]

There are two claims being made here. The first is that people will avow the norms they hold when they are unconstrained. This seems straightforward. The converse is also claimed: that if people avow something when unconstrained, then they accept it. This is not as straightforward. One might relax their position on a matter to entertain ideas which are contrary to the ones they usually hold to test the position's consistency. One might avow extreme fascism just to see if it is internally coherent, but that doesn't mean that they actually accept such a view. Furthermore, with this problem exposed, normative avowal alone ceases to be distinctly expressivist. Gibbard claims that what is distinctive about normative beliefs is their link to motivation and action, if that link is gone, Gibbard's explanation becomes no better than a basic descriptive one.

The last problem that I want to draw attention to is the difficulty Gibbard's theory has in distinguishing specifically moral judgments from other kinds of judgments. To recap, Gibbard's account claims that we know when an act or judgment is a moral one by the emotions it elicits: guilt on the part of the perpetrator, and justified anger on the part of onlookers. Aside from the lesser problem that these emotions are going to be difficult to identify because they look different in different people, we have a problem of judgmentalism.

We might say that moral judgments cause guilt and anger when one believes that they have done something morally wrong – thereby identifying the judgment as a moral one. This is

[65] Allan Gibbard, "Wise Choices, Apt Feelings," 74.

obviously wrong, and Gibbard recognizes this by saying that "I can feel angry at you and yet think that it makes no sense to do so; I can think that really you have acted as you should."[66] But if anger and guilt are not always readily recognized in different people then "the norm-expressivist account of moral judgments is circular: it invokes anger and guilt to characterize moral judgments, but we must understand moral judgments already if we are to characterize guilt and anger."[67] We need a non-judgmental way to identify guilt and anger.

Gibbard's solution is attributional theory: "When I think of myself as guilty, I see myself as being in a state that I conceive as follows: it is typically caused by my own acts of certain kinds, it is expressed by a guilty mien, and it typically moves me toward apology and amends."[68]

Thus, an act does not cause guilt unless I believe that I did a relevantly moral act. This might appear to be circular as well, but it only falls to that charge if moral judgments define relevant acts. To avoid further circularity, Gibbard says that an observer who wants to characterize guilt "can appeal directly to the central kind of circumstances in which, in my culture, one is thought to be at fault. Those circumstances will be part of the cluster the observer uses to define guilt."[69] Guilt then, is defined by cultural considerations which are then extended to similar cases.

There are a few problems with this explanation. The first is that moral 'rebels' might rebel against the prevailing moral views to reform the moral culture they reside in. A common

---

[66] Allan Gibbard, "Wise Choices, Apt Feelings," 130.

[67] Allan Gibbard, "Wise Choices, Apt Feelings," 148.

[68] Allan Gibbard, "Wise Choices, Apt Feelings," 148.

[69] Allan Gibbard, "Wise Choices, Apt Feelings,"149.

example of this behaviour in Western culture would be from vegetarians or vegans. Such people would feel bad if they ate meat, but would the emotion they feel be defined as guilt on Gibbard's account? Clearly, eating meat is not frowned upon by the larger population, and presumably, they do not feel the same sort of negative emotion that they might if they lied or betrayed a friend; furthermore, they wouldn't have someone to apologize to if they were to eat meat. By these standards, they don't fall into Gibbard's category of guilt, yet they still are feeling guilty by eating meat. If asked, they would report feeling guilty because they did something wrong by eating meat, but without reference to the moral judgment (that eating meat is wrong), it is difficult to see how Gibbard could classify their feelings as guilt on his theory, thus lapsing again into circularity.

II. Conclusion

I conclude that Gibbard's norm-expressivism fails to solve these particularly expressivist problems. Some of his solutions make progress, but end up needing to give up expressivism to do so. Gibbard claims that people are morally motivated when they accept systems of norms that permit and disallow certain moral actions (eliciting the emotions of guilt and anger). We have seen that, as described, Gibbard's theory becomes circular, needing to identify a moral judgment as such before seeing that the emotions of anger and guilt are warranted.

Chapter 3

NEO-EXPRESSIVISM AND COGNITIVE THEORIES

In the previous chapters, we looked at two internalist accounts of moral motivation with foundations in sentiments. Prinz's account was decidedly non-cognitive and Gibbard's (while non-cognitive) tried to straddle the line between cognitivism and non-cognitivism. But why favour non-cognitive accounts of moral motivation over cognitive ones? In this chapter, we will examine two views: the first critiques Gibbard's hybrid expressivism as not being truly hybrid and proposes a 'true' expressive view, and the second presents a cognitive account of moral motivation.

I. Neo-Expressivism

A. Critique of Expressivism

As we have seen with both emotivist and expressivist accounts, sentences containing ethical propositions communicate a specific emotion or desire located within a shared understanding of semantics; in more general terms, the meaning of ethical sentences is not about what they describe propositionally, but the non-cognitive attitudes they express. Difficulties arise when comparing the semantic form of ethical sentences with other declarative sentences. What justification do philosophers have in defending a cognitive understanding of non-ethical sentences and denying the same meaning-based interpretation of ethical sentences? Expressivists like Gibbard champion views which are hybrid in nature; that is, they have

developed explanations which express the meaning of ethical sentences by appealing to both cognitive and conative states.[70]

i. Ethical Sentences Embed Motivational States

In "(How) Is Ethical Neo-Expressivism a Hybrid View?", Bar-On et al. express confusion towards the efforts of expressivists, asking why it is that we should say that sentences – "ethical or not – are in the business of expressing mental state types (whether cognitive, conative-motivational, or some suitable hybrid)."[71] In their view, it is much easier to assume that sentences (regardless of subject matter) express propositions. To them, hybrid views seem to hybridize what expression is rather than the types of mental states expressed by an ethical sentence.

Instead of searching for a convoluted way of embedding motivation within semantics, why not grant that the propositions in ethical sentences do have real meaning, and that is the act of *making* an ethical claim that embeds mental states which indicate motivation?

The obvious advantage of such an approach is that it leaves philosophers free to take ethical sentences at face value: Declarative sentences of all types express their meaning through propositions. Why is this a desirable view to hold? First, the literal meaning of a sentence stays the same if that sentence is translated into another language. It would be silly to think that a sentence translated from English into French is a good translation if it expressed different propositional content after the translation. Secondly, in a declarative sentence, the same

[70] It is important to remember that even in Gibbard's view, expressivism is still the primary vehicle for understanding the semantic context of ethical sentences.

[71] Dorit Bar-On, Matthew Chrisman and Jim Sias, "(How) Is Ethical Neo-Expressivism a Hybrid View?," in *Having it Both Ways: Hybrid Theories and Modern Metaethics* ed. Guy Fletcher and Michael Ridge (Oxford: Oxford University Press, 2014), 223-247.

propositional content can be included with different syntactical structures without losing reference. The sentence "It is a nice day out" can make a simple statement of fact, but it can also appear within sentences like: "If it is a nice day out, I will go for a hike" or "It might be a nice day". Each sentence here has a different meaning, but clearly the reference of "nice day out" is the same between them.

Consider an ethical sentence we have previously looked at, "Bullying is wrong." Does this sentence also express propositional content? Syntactically, it is declarative, and it can create non-ethical sentences such as: "Bullying is common," and "Teachers try to stop bullying" with the same logical form. It is also clear that ethical sentences such as "Bullying is wrong", when translated into another language, carry the same propositional content, and would be considered a good translation. Furthermore, it seems that ethical propositions function identically to non-ethical propositions in different semantic contexts. I can say "Bullying is wrong," but also "If bullying is wrong, then it is wrong to teach your little brother how to bully," and "Bullying might be inappropriate." I might suspect that your son has been bullying, and you might doubt that your son has been bullying, in which case "Your son has been bullying" is a good representation of what I suspect and what you doubt. Just as normal sentences do, ethical sentences express the ability to communicate the same proposition across sentences with different syntax.

This shouldn't be surprising to anyone. It isn't by accident that people use ethical propositions in the same fashion as normal propositions in their sentences; nor is it because we do not know what to do with ethical propositions, simply shoehorning them into 'normal' syntax because we cannot think of a better way to utilize them. It simply seems to be the case that we talk about real things when we discuss ethics. If treating ethical sentences like normal sentences is as straightforward as claimed here, then why do expressivists try to re-work the whole

semantic use of ethics into something different? This seems to boil down to two 'problems' that expressivists feel the need to fix:

a) Avoiding 'spooky' irreducible normative facts

b) Explaining the strong link between ethical propositions and motivation

With regard to a), propositions are most often regarded as being bearers of truth, yet while it is most often possible to assess the truth conditions of many declarative sentences through observation, there does not appear to be a similar way of doing so for ethical claims.[72] Regarding b), expressivist theories claim to be able to explain the connection between the uttering of ethical propositions and the desire to act by embedding motivational attitudes into the meanings of ethical sentences themselves.

Bar-on et al. feel as though a propositional understanding natural to both moral and normal sentences can be maintained – remaining neutral (and allowing for anti-realism) on a). Regarding b) they hold that the best way to understand the connection between motivation and ethical semantics is not some quality in the meanings of such sentences, but rather that sentences express one thing (semantically) and that the speakers of these sentences express something else through the action of speaking the sentence (motivationally).[73]

[72] One can assess the truth conditions of the declarative sentence "Dave has brown hair" through observation. I can determine if Dave indeed has brown hair by finding Dave and observing the colour of his hair. There are many declarative sentences which we cannot check the truth value of because of our limited capabilities. "There are 3 little green men living on a planet in the Andromeda galaxy" has a truth value, but we don't have the ability to observe if it is true or not because of limitations in our ability to observe the planets in the Andromeda galaxy. Nevertheless, we do know how we *would* investigate the truth of this sentence if we had the ability to do so. There doesn't seem to be a similar method of knowing how to determine the proper truth conditions of ethical sentences.

[73] Bar-on et al., "(How) Is Ehtical Neo-Expressivism a Hybrid View?,".

ii. A-Expression and S-Expression

There are an enormous number of ways a person can express themselves, both verbally and non-verbally. A flinch, a jerking of the body, crying out, saying "ouch", stating "that hurts" are but a small number of ways in which a person expresses that they are experiencing pain. Some expressions seem to be universal – smiling and laughing seem to be expressions of happiness and humour around the world. Other expressions, such as hitting someone with the bottom of a shoe to insult them, or sticking out one's tongue are cultural expressions, learned from others in one's community. In addition to these, verbal expressions like, "This is the most fun I've had in a long time!" express our feelings as a matter of fact. Other verbal expressions, like "I promise" are speech *acts*, which express mental assertions. Whether verbal or non-verbal, expressions all communicate *states of mind*, as opposed to propositions, which express states of fact. Wilfred Sellars conceives of this kind of expression as a relation between the mental state of an individual, and the acts that individual does to express that mental state – "a-expression."[74]

A-expression is something intentional, and either verbal or non-verbal that expresses a feeling or intention. This is different from "s-expression", where semantic meanings are used to express propositional content. So when someone says "It is 25 degrees outside", they s-express the proposition that it is 25 degrees outside. Words like "truth" or "avarice", when spoken, s-express their meanings. When people hug each-other, or say "I enjoy spending time with you", they a-express their feelings towards each-other, while allowing that their spoken words express propositional meanings.

[74] Wilfrid Sellars, "Language as Thought and as Communication," *Philosophy and Phenomenological Research* 29, no. 4 (1969).

Bar-on finds it helpful to explain that when one a-expresses using a sentence, the sentence uttered retains its linguistic meaning.

> "It's great to see you" and "I'm so glad to see you" each have their own meaning, in virtue of the linguistic rules governing the lexical items and their respective modes of composition. Each s-expresses a certain proposition. What proposition? Well, setting aside some nuances about the context-sensitivity of indexicals, it's most natural to say that the former sentence expresses the proposition that it's great to see the addressee, and the latter sentence expresses the proposition that the speaker is happy to see her addressee. It is because they express propositions that these sentences can partake in logical inferences and stand in systematic logico-semantic relations to other sentences (and, in particular, can be embedded in negation, conditionals, intensional contexts, etc.). For all that, we think that normal cases of producing unembedded tokens of these sentence types, in speech or in thought, are cases of one's *directly expressing one's joy*.[75]

This illustrates the important contrast between expressing a state of mind and telling about states of mind. Anyone can say *that* Steve is feeling inspired to paint a picture, and they can do so truthfully and reliably. But it seems right to say that only Steve is able to properly *express* his feeling of inspiration – by spending his free time working on the painting, or telling his friends "I'm so excited to be working on this painting". To recap this using our understanding of Bar-on's terminology, we can say that when someone says "Steve is feeling inspired to paint", they use a sentence that s-expresses the proposition that Steve feels inspired to paint, and if they speak truthfully, then they a-express their *belief* that Steve is inspired. Steve's choice of activities a-expresses his inspiration to paint, and his uttering of the sentence "I'm so excited to be working on this painting" s-expresses that he is excited while at the same time a-expressing his excitement. If we allow for a-expression through sentences which s-express a proposition, we can explain the epistemic asymmetry between avowals and third-person reports

[75] Bar-on et al., "(How) Is Ethical Neo-Expressivism a Hybrid View?,".

of the same states without compromising the semantic continuity between avowals and other statements. The asymmetry described here is not between certain propositions, but between expressions of a mental state, and semantic reports of that state.

The understanding of a-expression and s-expression is what Bar-on et al. think explains the expressive nature of moral claims, and why motivation is so closely linked with them.[76] With an understanding of the distinction between a-expression and s-expression, one can hold to a motivational internalist position, and understand that moral sentences have the job of expressing the speaker's moral views, while understanding that there is a difference between this expression and the vehicle used in making it. "Bullying is wrong" utilizes a sentence as a vehicle to s-express a proposition that is either true or false.

B. Neo-Expressivism

The following highlights the difference between traditional neo-expressivism. On avowal neo-expressivism, when one makes an avowal, they are expressing the mental state that the sentence describes. On ethical neo-expressivism, ethical sentences s-express propositions which allows for semantic continuity with other sentences. The apparent asymmetry between these sentences and the motivation they provoke is explained by appeal to the expressive character of the *act* of making an ethical claim.

This view grants that expressivists have been correct in identifying an expressive element of ethical discourse, as it is part of the nature of ethical claims to express our motivational attitudes. It also avoids the pitfalls of cruder forms of expressivism, which requires that we

[76] Dorit Bar-on and Matthew Chrisman, "Ethical neo-expressivism," in *Oxford Studies in Metaethics, vol. 4*, ed. Russ Shafer-Landau (Oxford: Oxford University Press, 2009), 132-165.

attempt to cobble together semantics and emotion-rooted expression in awkward ways. When a person makes an ethical claim, a distinction can be observed between the act of making an ethical claim and the vehicle used to make it. Humeans might claim that cognitive states (such as beliefs) cannot motivate action on their own, and hold that when one makes an ethical claim, one should be motivated to act in accordance to that claim. The expressivist claims that when an ethical claim is made, the one who makes the claim a-expresses the mental states (usually emotional) that are essential to making the claim in the first place. A.J. Ayer recognized that this expressivist claim is best understood *without* supposing that the sentences used to make ethical claims are expressing propositions that self-ascribe those states.[77]

Bar-on's interpretation agrees with Ayer on this point. Ethical claims do reveal motivational attitudes, not because the claims themselves report them, but because ethical claims a-express those motivational attitudes directly. To summarize, Bar-on's ethical neo-expressivism claims that:

> (i) As a species of evaluative claims, ethical claims understood as acts are different from ordinary descriptive claims in that agents making them (in speech or in thought) a-express motivational attitudes.
>
> (ii) The vehicles used in making ethical claims – typically, ethical sentences – are semantically continuous with ordinary descriptive sentences in being truth- evaluable, embeddable in truth-functional as well as intensional contexts, logical inferences, etc. This is because they s-express true or false propositions.[78]

---

[77] A.J. Ayer, *Language, Truth and Logic* (New York: Dover Publications), 104-8.

[78] Bar-on et al., "(How) Is Ethical Neo-Expressivism a Hybrid View?,".

i. Adapted to Motivational Internalism

This evaluation of the nature of moral motivation can also be adapted to the many motivational internalists who find sentimentalist-based explanations of moral motivation (like that of Prinz and Gibbard) attractive. Internalists claim that if one makes a moral judgment such as "Lying is wrong", they are not competent moralizers unless they *feel* something towards lying, whereas one who says "Many people lie" can be indifferent to the spoken proposition.

Recall the distinction Bar-on makes between expressing a state of mind and telling *about* a state of mind. When Steve reports, "I'm feeling inspired to work on my painting", the hearer expects that Steve is currently feeling inspired, and is expressing his current feelings towards painting. The same expectation holds for any verbally-expressed mental state. When Steve expresses that he is inspired, or angry, or sad, he is not simply expressing the belief that he is these things. It may be conceptually possible for one to report one's mental state without feeling the requisite emotion, but such a state of affairs would be highly specialized and unusual – far outside the norms for such things. It can be said that it is a *propriety condition* to be in the self-described mental state when reporting on such things.

Bar-on's neo-expressivism regards the nature of ethical claims to be analogous to the reporting of mental states. When someone makes an ethical judgment (as opposed to reporting a fact related to ethical matters), they a-express the relevant motivational attitude – the attitude being the reason the person is motivated in the first place. Although it may be conceptually possible for one to make an ethical claim without having the relevant motivational attitude, having the relevant motivational attitude can be said to be a *propriety condition* on making genuine ethical judgments. A person making an ethical judgment, and does not possess the requisite motivational attitude has violated a propriety condition of making moral judgments, and

of course, to make a genuine moral judgment, one must be competent enough to grasp the propriety conditions required.[79] This is very similar to the point made in chapter 1 by Jesse Prinz – who stated that a person who makes a moral claim while feeling no sentiment, has done so defectively.

ii. Propriety Conditions

Propriety inform us of the norms required to do something *properly*. A propriety condition to making a promise is that the one making the promise intends to uphold their word. If I say "I promise…" without intending to do what I promise, then I have not properly made a promise. Does this mean that without the intention to follow through on my promise, I haven't made a promise at all? No, because I still have the understanding to know what a proper promise is and means despite my lack of intention to follow through on it. I have made a promise *incorrectly* in the same way that someone who burns the food they are cooking cooks *incorrectly*.

For ethical claims, we can similarly say that someone who is competent in ethical claims (understanding ethical vocabulary and the motivational attitudes required for making moral claims) might make a moral claim while failing to possess the requisite motivational attitude. Thus, when a moral claim is made without feeling motivation, it doesn't have to be the case that gibberish is being spoken, devoid of meaning. Rather, an ethical claim is being made, but incorrectly.

[79] For a comparison, think of slurs or cursing. We don't hold accountable a foreign speaker with no knowledge of English at fault for repeating the curse words she heard in a gangster rap song with no knowledge to their meaning. The words are spoken, but without satisfying the propriety conditions for cursing.

Pure expressivism has had difficulty in developing an account of meaning for ethical sentences since it was thought by expressivists that ethical sentences expressed non-cognitive attitudes. Understandably, expressivists had difficulty explaining why this state of affairs existed for ethical sentences alone, and so a larger project of expressivism was to explain that *all* sentences should be interpreted in light of the attitudes they expressed. More recently, hybrid expressivists moved back to a distinct focus on ethical sentences, claiming that only they express a distinct hybrid nature – consisting of both belief and desire-like elements. The belief-like attitude exists to explain the semantic similarity between ethical sentences and non-ethical sentences, and the desire-like attitude is essentially the expressivist concept that an attitude is being expressed.

iii. Summary

The view we've examined here claims to offer a neo-expressivism as opposed to Gibbard's hybrid expressivism, but is this neo-expressivist view just another form of hybridism? Bar-on claims that when a moral judgment is made, the speaker s-expresses a proposition while a-expressing the motivational attitude; which might seem to indicate that this is just a new way of affirming hybrid-expressivism. This does not seem to be the case however. Bar-on does not claim that a-expression and s-expression are two separate things that combine to provide meaning and context to ethical sentences. He also does not locate a-expression in a conventionally implicated proposition that self-ascribes the attitude.[80] Hybrid expressivism must use a non-conventional semantics for ethical sentences, followed by a merging with the mental

[80] See Copp (2001, 2009) and Finlay (2004, 2005). Bar-On and Chrisman 2009: 150-158 compares and contrasts ethical neo-expressivism with these implicature views.

state that ascribes the relevant attitude – hence, hybridism. Bar-on's neo-expressivism allows for conventional understanding of the semantics of ethical sentences, while allowing epistemic space for these sentences to express attitudes as well, in a way that remains neutral on ethical realism and cognitivism.

II. Cognitivism

Bar-on's neo-expressivism has some interesting answers to the problems faced by sentimentalist and hybrid theories. But why be relegated to the notion that expressivism is the best explanation for moral motivation? Laura and François Schroeter propose that one can be a pure cognitivist, and claim that the semantic function of normative predicates is to ascribe a property.[81] As we have seen, the common criticisms of ethical cognitivism tend to focus on the difficulty cognitivists have explaining how motivation arises particularly from ethical sentences while not for descriptive ones. Responding to this difficulty, Schroeter proposes a new cognitivist account of normative judgment grounded in a general model of concepts, where concepts are individuated similarly to how separate thoughts can be linked on a common topic.

A. *de jure* Sameness of Thought

Conceptually, there is a distinction at the level of thought between representing the same and representing *as* the same. You might see the same person on two separate occasions (briefly, in a crowd of people), registering the same visual details on both sightings. On the second sighting though, you may fail to recognize that you have seen that person beforehand – you don't

[81] Laura Schroeter and François Schroeter, "Why Go Hybrid? A cognitivist Alternative to Hybrid Theories of Normative Judgment," in *Having it Both Ways: Hybrid Theories and Modern Metaethics* ed. Guy Fletcher and Michael Ridge (Oxford: Oxford University Press, 2014), 248.

recognize it *as* the same.

By contrast, when you exercise in a chain of reasoning, your thoughts will automatically be presenting *as* applying to the same topic. I think of Russell Crowe, and then wonder how many movies he's been in, how many countries he's filmed in, what his net worth is, etc. From my perspective, it is plainly obvious that there is only one topic in question – Russell Crowe. Schroeter calls this *de jure* sameness of topic.[82] Not every thought pertaining to the same topic displays *de jure* sameness of topic. I may be surprised to discover that a disguised Russell Crowe played a part in a movie where I didn't recognize him. Although this is an identity judgment, it does not seem to me in the moment that *de jure* sameness of topic applies.

Frege recognized that *de jure* sameness of topic was essential to individuating the meanings of co-referential expressions.[83] This seems correct since learning would be impossible without having some sort of organizational structure to the thoughts we amass pertaining to every topic. As an aspiring chef, I need to be able to rely on my knowledge about how to prepare certain foods, form new beliefs about those foods as my knowledge increases, stay on topic while experimenting with new ingredients etc. This process depends on the organization of my thoughts to pertain to the thing in question. Schroeter thinks that it requires something like a 'mental filing system' "for binding and segregating attitudes by topic, in such a way that you're immediately disposed to take bound attitudes as obviously and incontrovertibly pertaining to the

[82] The notion of *de jure* sameness of topic was introduced in (Schroeter 2007, 2008). For further discussion of its epistemic role, and its importance to concept individuation, see (Schroeter 2012).

[83] Gottlob Frege, "On Sense and Reference," in *Translations from the Philosophical Writings of Gottlob Frege*, trans. M. Black (Oxford: Blackwell, 1892).

same topic."[84]

Using 'filing system' terminology is a helpful metaphor because it helps get across the functional concept being deployed to describe *de jure* sameness.[85] In an office, a filing cabinet serves the purpose of organizing documents to pertain to the same topic, and differentiate it from the other topics contained in the filing cabinet. When you pick a file out of the cabinet labeled "John Smith", you immediately expect that all the content within pertains to the individual John Smith and not something else. The thing to note here is that you form this expectation based on the label on the file, not because you went through each document and arrived at the conclusion that all the documents together pertained to John Smith.

Schroeter takes this concept as applicable to thoughts. In order to have a conversation about anything, or make an argument, or even pay attention to the plot of a movie, one's thoughts need to be organized as pertaining to the topic in question. This can happen without active concentration on a topic as well. I might see someone walking a dog, and immediately be reminded of several relevant facts applying to dogs that I was not consciously thinking of beforehand – what kinds of breeds are my favourite, how much backyard space is best for a dog etc. For me to be able to do this, there must be some kind of mental structure that organizes thoughts on a topic for quick access.

B. Normative Beliefs

Normative beliefs have this sort of conceptual structure as well. To judge a moral action

[84] Laura Schroeter and François Schroeter, "Why Go Hybrid? A cognitivist Alternative to Hybrid Theories of Normative Judgment," 249.

[85] Early proponents of the view that cognition of individuals involves mental files or 'dossiers' of information about a topic include (Grice 1969; Evans 1973; Strawson 1974, ch. 2).

as right or wrong, we must be able to track the concepts of moral rightness and wrongness. An account of normative thought that cannot keep track of *de jure* sameness of topic fails automatically. Jesse Prinz claims that "to believe something is wrong in a non-deferential way is to have a sentiment of disapprobation toward it"[86] On Schroeter's analysis, this is a category mistake.[87] A sentiment doesn't bind moral attitudes together in such a way that they all appear to *de jure* pertain to the same topic. Any theory that proposes to reduce normative judgments to attitudes or sentiments will fail to explain how the attitude or sentiment creates *de jure* sameness of topic through our thoughts. Pure and hybrid expressivism conflict with this as well. Recall that Gibbard's hybrid-expressivism claims that when one judges something as rational, they accept norms that permit it.[88] This causes it to fall into the same problems that Prinz's sentimentalism does - different decisions or states of norm-acceptance aren't bound in such a way that they strike the subject as pertaining *de jure* to the same topic.

> Cognitivists in the metaethical literature have tended to highlight the representational (mind-to-world) function of concepts: concepts constitute ways of picking out objects, kinds, or properties in thought. But your concept of MORAL WRONGNESS doesn't merely serve to demarcate the property of being wrong, it also serves to cross-index your various MORAL WRONGNESS attitudes as pertaining *de jure* to the same topic. This mind-to-mind aspect of concepts – the way they mark apparent sameness of topic among mental states – has been insufficiently emphasized in the metaethical literature. But it's crucial to understanding conceptually articulated thought.[89]

We can see this mind-to-mind concept more clearly by examining the logical concepts of

---

[86] Jesse Prinz, *The Emotional Construction of Morals* (Oxford: Oxford University Press, 2007), 94.

[87] Schroeter's argument doesn't address the question of whether or not emotions 'track' features of the world. According to Prinz, emotions like fear represent anthropocentric properties like danger. Schroeter, however, is focussing on a mind/mind relation rather than a mind/world relation where *de jure* sameness of topics is the thing that links propositional attitudes.

[88] Gibbard, "Wise Choices, Apt Feelings," 46.

[89] Schroeter, "Why Go Hybrid? A cognitivist Alternative to Hybrid Theories of Normative Judgment," 252.

validity, contradiction, and consistency:

1. The morning star is visible.
2. Therefore, the evening star is visible.

The above is not logically valid, but at the same time, we know that the truth of the first premise guarantees the truth of the conclusion since the morning and evening stars are in fact the same object. Logical validity requires more than *de facto* sameness of topic; it also requires *de jure* sameness of topic since it is possible that the speaker does not realize that the morning and evening stars are the same object.

C. Concept Identity and Reference

To proceed, Schroeter claims the need for a theory of concept identity and reference. How is it that both a belief and a desire refer to the same concept? And secondly, how is the *reference* of thought determined? What makes it the case that my 'John Smith' thoughts refer to John Smith, my 'dog' thoughts refer to dogs, and my 'moral wrongness' thoughts refer to moral wrongness?

Frege's traditional account claims that a single concept explains both roles. For *concept identity,* it can be said that two thoughts refer to the same concept if each thought is associated with the same reference-fixing criterion – cognitive dispositions that put the subject in a position to recognize the reference. For *reference,* it is a token thought determined by the reference-fixing criterion associated with that thought in context. On this traditional account of Frege's concepts are finer-grained than reference, since different criteria can pick out the same object, kind, or property. Because of this "morning star" and "evening star" deploy different concepts.

Frege's account can also explain why different thoughts about the same topic seem obviously co-referential. The reason you cannot imagine a case where the morning star is not the evening star, is because there is a precise match in the reference-fixing criteria between both objects.

Schroeter favours a 'binding' account that separates concept identity and reference. Firstly, what makes it the case that two token thoughts deploy the same concept is a relational fact common to both thoughts. Secondly, isolated facts do not determine reference about that token, but rather by facts about a larger group of facts to which that token belongs. On Schroeter's binding account, there is no need for each token to be associated with the same reference-fixing criterion to guarantee co-reference.[90]

Schroeter's view of concept identity is like anaphora in normal language. Consider:

A: "Papa Murphy's is a great restaurant."

B: "I've never heard of it."

In this conversation, it is obvious to both parties that they are talking about the same thing – Papa Murphy's. But clearly the level of understanding relating to this topic is not shared. A clearly seems to have eaten at Papa Murphy's and has an understanding of the menu, such that they feel justified in calling the restaurant 'great'. B, on the other hand seems to know nothing about Papa Murphy's other than the fact that is a restaurant (just learned by hearing A's

[90] An intuitive way to get an idea for how Schroeter's binding account works is to consider how personal identity works. What makes it the case that the baby born in 1986 is the same person as the one currently writing this paper? Traditional accounts of personal identity try to refer to some feature of the person at each temporal stage that is necessary and sufficient for making that person the same person. This could be a soul, personality traits or some other metaphysical element. Relational accounts of personal identity refer to causal relations – linking memories or experiences. Schroeter's account of concept identity is similar to this relational account of personal identity.

comment). Clearly, the *de jure* sameness at play in this exchange is not based on a prior substantive understanding of both parties about the concept in question. All that is needed for *de jure* sameness of topic here is that the pre-reflective linguistic mechanisms of discourse establish a direct cognitive binding so that both parties realize that they are talking about the same topic.

Of course, in addition to apparent *de jure* sameness, bounds thoughts also must *actually* be representing the same topic. Consider the continuation of our anaphora example:

A: "Papa Murphy's is a great restaurant."

B: "I've never heard of it."

C: "I've been there once, I didn't think it was that good."

D: "I know where that is, at the end of 19th St."

This exchange follows the same pattern established above that creates the appearance of *de jure* sameness of topic. In this exchange the first three speakers co-refer, despite having a large variation of knowledge and opinion towards the topic. The fourth speaker however, does not – it happens to be the case that the restaurant is nowhere near 19th St., and the speaker has confused Papa Murphy's with another restaurant. It is important to note that despite the appearance of *de jure* sameness in anaphoric relations, the binding account is a non-starter if in fact speakers do not *de facto* refer to the same concept.

Bringing this back to Schroeter's mental filing system, the speaker's mental filing system create a speaker's relevant epistemic connections to a given topic. For any topic the subject has knowledge of, reflecting on that topic will bring up a host of thoughts, attitudes, and dispositions

which all seem to the subject to be obviously and incontrovertibly related to the concept being thought of. Mental filing systems cannot be shared between individuals, but as the anaphora example shows, the appearance of *de jure* sameness can be generated through conversation, and so the mental linking of concepts must occur through linking thoughts and dispositions to certain words or concepts.[91] These interpersonal binding mechanisms are not as reliable as personal ones as there is a higher chance that we fail to co-refer when I use names or concepts that are not as familiar to the listener.

Frege's traditional account of concept identity states that individual concepts determine semantic value in isolation. Thus, reference-determination will be explanatorily prior to apparent *de jure* sameness relations – the two tokens have the appearance of *de jure* sameness because they are independently linked to the same reference-fixing criterion.

Schroeter's binding model reverses this explanation: basic units of interpretation are not created in isolation, but in relation to the entire host of elements that the subject takes as being *de jure* related to the concept in question. Thus, on the binding model, apparent *de jure* sameness is explanatorily prior to reference-determination.

D. A Cognitive Account of Motivation

We now return to the question of how the binding model of concepts benefits a pure cognitivist account of moral motivation. If (such as is proposed on the traditional Fregean model) normative judgments achieve motivational qualities in virtue of simply attributing a property considered in isolation, how could we understand it other than by assuming that judgment and

[91] Sam Cumming, "Creatures of Darkness," *Analytic Philosophy* 54, no. 4 (2013): 379-400.

motivation are external and contingent? It would be like justifying the fact that most Canadians like maple syrup because many Canadians do. On this reasoning, the connection between judging an action to be right, and then feeling motivated to act on it would be justified because most people do act that way.

The simplest solution to this problem is strict motivational internalism, where a subject is not a competent moralizer unless she is also appropriately motivated. As we have seen by looking at motivational internalism in the previous chapters, there are several challenges to holding this view. Schroeter's binding model has made it clear that motivation is not essential for conceptual competence in ethics, but it does not follow that motivation is semantically irrelevant. Instead, motivation takes on the important role of fixing reference and determining competence at the interpersonal level.

On the binding model, semantic use of a concept communally is determined through reflective equilibrium linked with the socially extended representational tradition to which it is linked. Correct semantic interpretation occurs when the concept captures what is most important to sustaining a shared tradition. In practice, this means that we are concerned with normative concepts such as 'moral rightness' because communally, people cite 'moral rightness' as justification for actions and positions. Because of its centrality to representational practice, motivation has a significant role in guiding correct interpretation of moral (and other normative) concepts. Specifically, motivation determines which property is assigned as the conceptual reference.

At the broadly normative level, determining the point of a representational tradition is important for how we understand the topic in question. For example, a debunking account of

morality criticizes the idea that moral systems privilege one group of people over another – thus making the term 'morally right' refer to the system of norms that benefit the privileged group.

Similarly, motivation is important for determining which reference is assigned to the moral concept. Take three similar yet distinct normative concepts like 'wrong', 'bad', and 'undesirable'. For 'wrong', proper communal competence is important – it provides a reference for how to evaluate and categorize actions. Because of this, the reference fixing quality to 'wrong' must be action-guided, and therefore, motivation is important because it provides a direct guide to action. By contrast, categorizing something as 'bad' doesn't carry the same action-guidedness. 'Bad' is a gradable property – papers receive higher or lower marks because of this quality. This is not the case for 'wrong'. The action-guiding role of normative concepts helps assign a categorical property as the reference rather than a gradable property. Additionally, the use of motivation to fix moral normative concepts brings the reference into the sphere of human ability in a way that gradable concepts cannot. 'Good' as a gradable concept can easily outstrip the human capacity for good – as is seen when attempting to comprehend the infinite goodness of God. Normative categorical judgments of right and wrong operate in such a way that we do not expect them to outstrip human capability. When we judge an action as morally wrong, we tend to use the term with the implication that the agent could have feasibly acted otherwise. The point here is that motivation need not be semantically irrelevant on the binding model as it is how reference is fixed for normative concepts.

In regards to competence, the binding model states that keeping track of sameness communally through apparent *de jure* sameness is required. Thus, even though psychopaths do not feel motivation to act on their normative judgments, it does not follow that they are not competent with moral normative concepts. How then does motivation factor into moral

competence?

As we saw in chapter 1, some philosophers have suggested that psychopaths simply use moral terms and concepts in a 'parasitic' way that imitates normal use of moral concepts without actually being moral themselves.[92] Clearly, further explanation is needed – individuals cannot be regarded as parasitically using concepts because they are abnormal with respect to the majority population. However, given the particular moral normative states under examination here on the binding model, competence must be linked to relevant concepts.

First, as we have seen, individuals are competent with normative concepts when they appropriately share a representational tradition with the larger community. Motivation is essential to fixing reference for these normative concepts, and therefore indirectly affects the tradition as a whole, regardless of whether or not a participant is cognizant of it. Thus, if psychopaths are linked to our 'normal' tradition of moral rightness or wrongness (of which motivation is central) then they can be competent.

The role that motivation plays in shaping a shared representational tradition of moral rightness and wrongness plays a constitutive role either directly or indirectly in an individual's competence with those concepts. Indirectly, motivation is constitutive of *everyone's* competence with moral concepts, since it plays an important role in fixing shared representational tradition. Directly, motivation is constitutive of *normal* users of the representational tradition since this is part of how the tradition is fixed in the first place. Thus, the binding model explains how motivation is constitutive of conceptual competence without imposing strict preconditions on its

[92] See: (Bedke 2009, 191-95; Blackburn 1998, 59–68; Dreier 1990, 9–14; Tresan 2006, 149–52; van Roojen 2010).

use.

In summary of Schroeter's binding model, they argue that pure cognitivism can explain the tricky phenomena that are usually cited in favour of hybrid theories. Motivation is not a strict necessary condition for moral competence, while not becoming irrelevant. Moreover, motivational expectations we have of moralizers is justified on the basis of normal conceptual competence. The link between motivation and normative judgment happens at the level of thought rather than language, as is the case with hybrid theories, and is anchored in the shared representational tradition of normative concepts.

III. Conclusion

We started this thesis with the goal of evaluating different accounts of moral motivation to see what best explains the uniquely motivating aspect moral judgments seem to possess. In the first chapter, we looked at Jesse Prinz's sentimentalist account. He claimed that emotions are both necessary and sufficient for moral motivation. Prinz claimed they were necessary due to the claim that psychopaths (who are emotionally deficient) do not understand the difference between moral and conventional wrongs. He claimed sufficiency on the basis of experiments done which showed that people in an emotionally negative state tend to judge moral situations more harshly than people who are not in the same negative state. These claims, while strongly rooted in intuition, failed to hold up under scrutiny.

In the second chapter, we examined Allan Gibbard's norm-expressivism. This account of motivation also has its foundation in sentimentalism, but directs the sentiments through a rational filter of what it 'makes sense' for the subject to do – guided by culture and social norms. We saw that Gibbard's norm-expressivism has to give up important aspects of expressivism to

remain coherent in addition to the fact that his theory becomes circular when put into practice.

In the third chapter, saw a response to expressivist theories by Dorit Bar-On, who showed that moral judgments can make sense of their cognitive language and motivational qualities though the distinction between a-expression and s-expression. His neo-expressivism avoids the problems of Gibbard's norm expressivism by separating the two forms of expression. Lastly, we looked at Schroeter's model of moral motivation based on pure cognitivism, where they argued that one need not be an expressivist to explain the motivational quality of moral judgments.

It is my opinion that the views examined in this last chapter hold the most promise for providing a robust account of moral motivation because of the way they can address the traditional problems of non-cognitive theories while leaving room for cognitivism.

Bibliography

Aharoni, Eyal Walter Sinnott-Armstrong, and Kent A. Kiehl. "Can Psychopathic Offenders Discern Moral Wrongs? A New Look at the Moral/Conventional Distinction." *Journal of Abnormal Psychology* 121 (2012): 484-497.

Ayer, A.J. *Language, Truth and Logic.* New York: Dover Publications.

Bar-on, Dorit, and Matthew Chrisman. "Ethical neo-expressivism." in *Oxford Studies in Metaethics, vol. 4*, edited by Russ Shafer-Landau. Oxford: Oxford University Press, 2009.

Bar-On, Dorit, Matthew Chrisman and Jim Sias, "(How) Is Ethical Neo-Expressivism a Hybrid View?." in *Having it Both Ways: Hybrid Theories and Modern Metaethics* edited by Guy Fletcher and Michael Ridge. Oxford: Oxford University Press, 2014.

Bedke, Matthew S. "Moral judgment purposivism: Saving internalism from amoralism." *Philosophical Studies* 144 (2009): 189-209.

Blackburn, Simon. *Ruling Passions: A Theory of Practical Reasoning.* Oxford: Oxford University Press, 1998.

Blair, R.J. "A cognitive developmental approach to mortality: investigating the psychopath." *Cognition* 57 (1995): 1-29.

Blair, R.J. et al. "Is the psychopath 'morally insane'?." *Personality and Individual Differences* 19 (1995): 741-752.

Blair, R.J. et al. "A Selective Impairment in the Processing of Sad and Fearful Expressions in Children With Psychopathic Tendancies," *Journal of Abnormal Child Psychology* 29 (2001): 491-498.

Blass, Thomas. *The Man Who Shocked the World: The Life and Legacy of Stanley Milgram.* New York: Basic Books, 2004.

Cleckley, Hervey M. *The Mask of Sanity: An Attempt to Clarify Some Issues About the So Called Psychopathic Personality.* Augusta: C.V. Mosby Co., 1988.

Copp, David. "Against internalism about reasons—Gert's rational options." *Philosophy and Phenomenological Research* 62 (2001): 455-461.

Copp, David. "Realist-expressivism and conventional implicature." In *Oxford Studies in Metaethics,* edited by Russ Shafer-Landau, 167-202. Oxford: Oxford University Press, 2009.

Cumming, Sam. "Creatures of Darkness." *Analytic Philosophy* 54 (2013): 379-400.

Davidson, Donald. “Causal Relations.” *The Journal of Philosophy* 64 (1967): 691-703.

Dawel, A et al. “Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy.” *Neuroscience and Biobehavioral Reviews* 36 (2012): 2288-2304.

Decety, Jean and Laurie R. Skelly. “The Neural Underpinnings of the Experience of Empathy: Lessons for Psychopathy,” in *The Oxford Handbook of Cognitive Neuroscience: Volume 2: The Cutting Edges*, ed. Kevin N. Ochsner and Stephen Kosslyn New York: Oxford University Press, 2013.

Dreier, James. “Internalism and speaker relativism.” *Ethics* 101 (1990): 6-26.

Frege, Gottlob. “On Sense and Reference.” in *Translations from the Philosophical Writings of Gottlob Frege*, trans. M. Black (Oxford: Blackwell, 1892).

Gibbard, Allan. *Wise Choices, Apt Feelings.* Cambridge: Harvard University Press, 1990.

Gilligan, Carol. *In a different voice: psychological theory and women’s development*. Cambridge: Harvard University Press, 1982.

Greene, Joshua et al. “An fMRI investigation of emotional engagement in moral judgment.” *Science* 293 (2001): 2105-2107.

Grice, H.P. “Utterer's Meaning and Intention.” *The Philosophical Review* 78 (1969): 147-177.

Haidt, Jonathan. “The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgement.” *Psychological Review* 108 (2001): 814-834.

Hare, R.M. *Essays in ethical theory.* Oxford: Oxford University Press, 1989.

Hare, R.M. *Freedom and Reason.* Oxford: Clarendon Press, 1963.

Hare, R.M. “Imperative Sentences.” *Mind* 58 (1949): 21-39.

Hare, R.M. *The language of morals.* Oxford: Clarendon Press, 1952.

Jones, Karen. “Metaethics and emotions research: A response to Prinz.” *Philosophical Explorations* 9 (2006): 45-53.

Kohlberg, Lawrence. “The Claim to Moral Adequacy of a Highest Stage of Moral Judgment.” *Journal of Philosophy* 70 (1973): 630-646.

Prinz, Jesse. “The Emotional Basis of Moral Judgments.” *Philosophical Explorations* (2006): 29-43.

Mackie, J.D. *Ethics: Inventing Right and Wrong*. London: Pelican Books, 1977.

Marturano, Antonio. "Non-Cognitivism in Ethics." *Internet Encyclopedia of Philosophy* http://www.iep.utm.edu/non-cogn/#H3.

Meffert, Harma et al. "Reduced spontaneous but relatively normal deliberate vicarious representations in psychopathy." *Brain* 136 (2013): 2550-2562.

Moll, Jorge, Ricardo de Oliveira, and Paul J. Eslinger. "Morals and the human brain: a working model." *NeuroReport* 14 (2003): 299-305.

Sanfey, Alan G, et al. "The Neural Basis of Economic Decision-Making in the Ultimatum Game." *Science* 300 (2003): 1755-1758.

Sauer, Hanno. "Psychopaths and Filthy Desks: Are Emotions Necessary and Sufficient for Moral Judgment?." *Ethical Theory and Moral Practice* 15 (2012): 95-115.

Schnall, Simone et al. "Disgust as embodied moral judgment" *Personality and Social Psychology Bulletin* 34 (2008): 1096-1109.

Schroeter, Laura. "Bootstrapping our way to samesaying." *Synthese* 189 (2012): 177-197.

Schroeter, Laura and François Schroeter. "Why Go Hybrid? A cognitivist Alternative to Hybrid Theories of Normative Judgment." in *Having it Both Ways: Hybrid Theories and Modern Metaethics,* edited by Guy Fletcher and Michael Ridge 248-273. Oxford: Oxford University Press, 2014.

Sellars, Wilfrid. "Language as Thought and as Communication." *Philosophy and Phenomenological Research* 29 (1969): 506-527.

Sias, James. "Ethical Expressivism." *Internet Encyclopedia of Philosophy* http://www.iep.utm.edu/eth-expr/.

Sinnott-Armstrong, Walter. "Some problems for Gibbard's norm-expressivism." *Philosophical Studies* 69 (1993): 297-313.

Stanford Encyclopedia of Philosophy. "Moral Anti-Realism." Accessed July 20, 2015. http://plato.stanford.edu/entries/moral-anti-realism/.

Stanford Encyclopedia of Philosophy. "Moral Motivation." Accessed July 20, 2015. http://plato.stanford.edu/entires/moral-motivation/.

Tresan, Joe. "De Dicto Internalist Cognitivism." *Nous* 40 (2006): 143-165.

van Roojen, Mark. "Moral rationalism and rational amoralism." *Ethics* 120 (2010): 495-525.

Wheatley, Thalia and Jonathan Haidt. "Hypnotic Disgust Makes Moral Judgments More Severe." *Psychological Science* 16 (2005): 780-784.